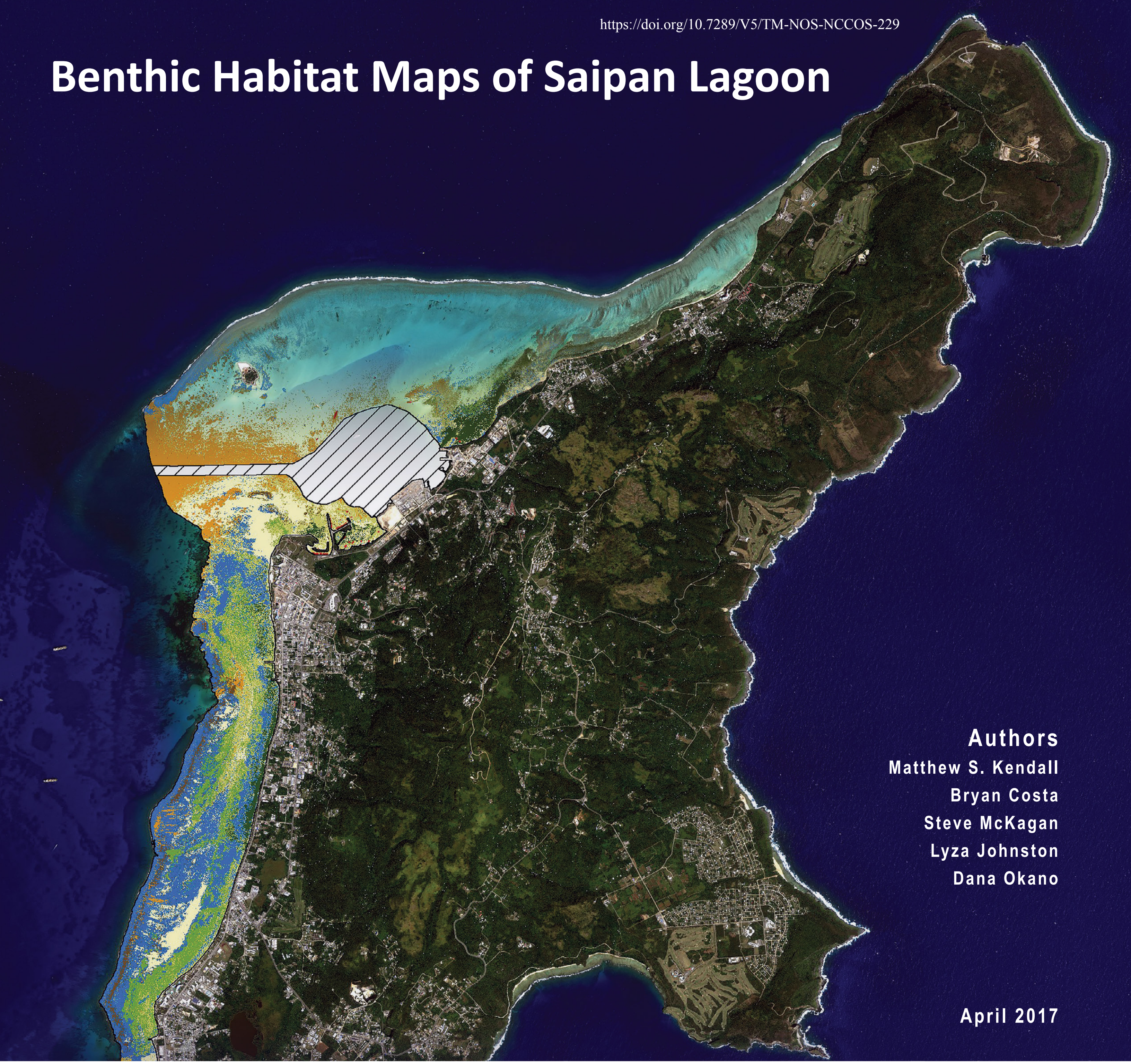
https://doi.org/10.7289/V5/TM-NOS-NCCOS-229

Benthic Habitat Maps of Saipan Lagoon



Authors

Matthew S. Kendall

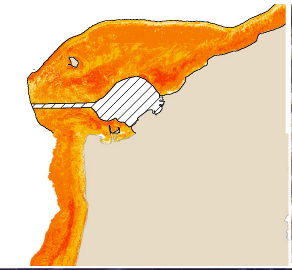
Bryan Costa

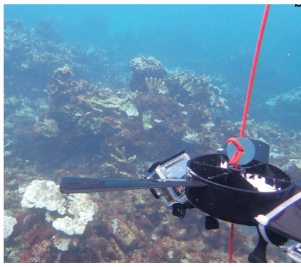
Steve McKagan

Lyza Johnston

Dana Okano

April 2017





NOAA TECHNICAL MEMORANDUM NOS NCCOS 229

NOAA NCCOS Center for Coastal Monitoring and Assessment

Suggested Citation

Kendall, M.S., B. Costa, S. McKagan, L. Johnston, and D. Okano. 2017. Benthic Habitat Maps of Saipan Lagoon. NOAA Technical Memorandum NOS NCCOS 229. Silver Spring, MD. 77 pp. https://doi.org/10.7289/V5/TM-NOS-NCCOS-229

Acknowledgments

This Saipan Lagoon mapping project was designed and executed by NOAA's National Centers for Coastal Ocean Science (NCCOS) in consultation with staff from the Common Wealth of the Northern Mariana Island's (CNMI) Bureau of Environmental and Coastal Quality (BECQ), and NOAA's Pacific Island Regional Office (PIRO) in the CNMI. The work was funded by NOAA's Coral Reef Conservation Program (CRCP) and NCCOS with in-kind contributions from BECQ, PIRO, and the University of Guam. Local experts were consulted before, during, and after the mapping process in order to receive suggestions for customizing the products to meet their needs. Briefings and other meetings included Fran Castro (Director, DCRM/BECQ), Ariele Baker (Coastal Hazards Specialist DCRM/BECQ), Rodney Camacho (BECQ), Robbie Greene (GIS Specialist DCRM/BECQ), and others. We collected and interpreted field data in collaboration with staff from the NOAA's CNMI Field Office (S. McKagan, D. Okano, and the 2016 NOAA Summer interns Katelynn Reyes and Aldebert Guerrero) and the CNMI Bureau of Environmental and Coastal Quality (L. Johnston, J. Iguel, R. Greene). Captain Frank Ramon from B-Sea Inc. shared his knowledge of the Lagoon and safely navigated us among field sites. Dr. Peter Houk from University of Guam generously provided data and insights from his previous mapping experience and field work in Saipan Lagoon and provided helpful feedback at multiple stages of the project. Rachel Husted (NOAA NCCOS) processed the many videos and photographs of field sites. Ken Buja (NOAA NCCOS) created the Biomapper data viewer. Tim Battista (NOAA NCCOS) orchestrated the WV2 satellite collection. We also thank John Christensen (NOAA NCCOS), Robbie Greene (BECQ), Anne Kitchell (Horsley Witten Group), and Laura Kracker (NOAA NCCOS) for their constructive review comments on a draft of this document.

The covers for this document were designed by Gini Kennedy (NOAA). The layout and formatting were completed by Sarah Hile (NOAA and CSS, Inc.). All cover photos were taken by M. Kendall and B. Costa. Editorial content was improved by Kevin McMahon (NOAA). CSS employees were supported under NOAA Contract No. EA133C-14-NC-1384 with CSS, Inc..

For more information on NOAA's National Centers of Coastal Ocean Science, please visit:
https://coastalscience.noaa.gov/

For more information on this project, please visit:
https://coastalscience.noaa.gov/projects/detail?key=271

Or direct questions and comments to:
Matthew S. Kendall, Ph.D.
NOAA/NOS/National Centers for Coastal Ocean Science
Center for Coastal Monitoring and Assessment/Biogeography Branch
1305 East West Highway, SSMC4, N/SCI-1
Silver Spring, MD 20910

240-533-0314
matt.kendall@noaa.gov

or

Bryan Costa
bryan.costa@noaa.gov

https://doi.org/10.7289/V5/TM-NOS-NCCOS-229

Benthic Habitat Maps of Saipan Lagoon

Prepared by:
Biogeography Branch
Center for Coastal Monitoring and Assessment (CCMA)
NOAA National Centers for Coastal Ocean Science (NCCOS)
Silver Spring, MD
USA

April 2017

Authors

Matthew S. Kendall[1], Bryan Costa[1], Steven McKagan[2], Lyza Johnston[3], and Dana Okano[4]

[1] *NOAA National Ocean Service/National Centers for Coastal Ocean Science, Biogeography Branch, Silver Spring, MD, U.S.A.*
[2] *NOAA National Marine Fisheries Service/Pacific Island Regional Office/Habitat Conservation Division, NOAA Field Office, Saipan, CNMI*
[3] *Bureau of Environmental and Coastal Quality, Division of Coastal Resources Management, Saipan, CNMI*
[4] *NOAA National Ocean Service/Coral Reef Conservation Program, Saipan, CNMI*

NOAA Technical Memorandum NOS NCCOS 229

United States Department of Commerce	National Oceanic and Atmospheric Administration	National Ocean Service
Wilbur L. Ross, Jr. Secretary	Benjamin Friedman Administrator, Acting	Russell Callender Assistant Administrator

*Staghon coral (*Acropora species*) in Saipan Lagoon. Credit: M. Kendall and B. Costa.*

Executive Summary

Saipan Lagoon plays a crucial role in the economy, tourism, and culture of the Commonwealth of the Northern Mariana Islands (CNMI). Understanding the spatial distribution and extent of Lagoon habitats is needed for planning sustainable economic development, evaluating zoning scenarios, minimizing user conflicts, ensuring public safety, and preventing environmental degradation in the Lagoon. To meet this need, NOAA's National Centers for Coastal Ocean Science (NCCOS) collaborated with CNMI's Bureau of Environmental and Coastal Quality (BECQ), and NOAA's Pacific Islands Regional Office (PIRO), to develop detailed maps of the distribution of seafloor habitats within the Lagoon and to assess changes that have occurred since the last maps were created based on 2001 satellite imagery.

Lagoon habitats were mapped using a WorldView 2 (WV2) satellite image collected February 5, 2016. Two types of map products were created. The first type describes the spatial distribution of five substrate types (i.e., 'Sand', 'Rubble', 'Pavement', 'Live Coral Reef (all spp.)', 'Upright Dead Coral Reef') and seven cover types (i.e., 'Bare Sand', 'Mixed Algae', 'Seagrass (*Enhalus acoroides*)', 'Seagrass (*Halodule uninervis*)', 'Live Coral (*Isopora palifera*)', 'Live Coral (Staghorn *Acropora*)', and 'Live Coral (other spp.)'. These classes were used to create 12 map layers, where 2 x 2 m grid cells in the map denote the probability that a given substrate or cover type is present (0 to 100%). The second product was a single-layer map based on a composite of the seven most common combinations of substrate and cover from field surveys plus 'Unknown' (for deep turbid areas) and 'Artificial' (for man-made objects). Habitat classifications assigned to the map grid in the composite map were 'Bare Sand', 'Seagrass (*E. acoroides*) on Sand', 'Seagrass (*H. uninervis*) on Sand', 'Mixed Algae and Seagrass on Sand', 'Pavement Colonized with Mixed Algae', 'Coral Rubble Colonized with Mixed Algae', and 'Live or Upright Dead Coral Colonized with Mixed Algae'.

Both map types were created using a combination of underwater videos and field data from 292 ground validation sites (collected July - August 2016), environmental predictor variables, and mathematical modeling. The 12 probability-of-occurrence maps were produced using Boosted Regression Tree models and twenty-eight environmental predictors (i.e., spectral bands from satellite imagery, geographic variables, and topographic surfaces). The composite Lagoon map was produced using Boosted Classification Tree models and the 12 probability-of-occurrence maps as predictors. Performance and accuracy of all models were evaluated using an independent set of field data from 273 accuracy assessment sites. Results indicate that substrate and cover predictions were robust, since they had little bias ($\overline{x}$ = 2% ± 1% SE) and explained over a third of the variation in the data ($\overline{x}$ percent deviance explained = 33.0% ±4.8 SE). Thematic accuracy of the composite maps was 86% correct overall, with user's accuracy of individual habitat classes between 80% and 100% correct.

Over 27 km^2 of seafloor was mapped in Saipan Lagoon. The extent of locations with high probability of occurrence for 'Live Coral Reef (all spp.)' was very constrained. The model predicted the highest likelihood of presence along the reef crest northeast of the harbor. The 'Upright Dead Coral Reef' model predicted likely presence along most reef crests and the sides of the harbor channel. Presence of 'Coral Rubble' was predicted on top and just inshore of the defined reef crest. The 'Pavement' model showed a similar pattern on reef crests, and also some areas where the Lagoon narrows at its northern and southern extremities. 'Sand' presence had the most widespread extent, with only the reef crest and immediate back reef areas showing low values. 'Mixed Algae' was the most widespread cover type and had high likelihood of presence throughout the Lagoon, except for the sand area east of Mañagaha. 'Bare' bottom was also widespread, except in those areas predicted to have seagrass and on reef crests. High probability-of-occurrence values for the individual species models were very constrained. Highest predictions for the seagrass, 'Seagrass (*E. acoroides*)', were close to shore near Tanapag, Smiling Cove Marina, and along Red Beach. Predictions for 'Seagrass (*H. uninervis*)' presence were high close to shore and just offshore of *E. acoroides*, around Tanapag, and spread throughout much of the southern Lagoon.

Executive Summary

The composite habitat map displays commonly co-occurring bottom types throughout the Lagoon in a single layer. 'Bare Sand' was the most abundant (32.6% of the Lagoon). The largest, continuous patch of 'Bare Sand' was north of the Harbor area. 'Pavement Colonized with Mixed Algae' was the next most abundant (25.7% of the Lagoon). Reef crests, backreefs, and the Lagoon floors approximately 1 km north and south of the Sugar Dock were dominated by this class. 'Live or Upright Dead Coral Colonized with Mixed Algae' was the third most abundant (13.9%). It was present in continuous patches along the channel and along the backreef of the northern Lagoon. 'Mixed Algae and Seagrass on Sand' was next (13.8%), however, its distribution was more diffuse than other habitats. 'Seagrass (*H. uninervis*) on Sand' was the fifth most abundant habitat (7.1%). Large meadows were most common between Garapan and Sugar Dock. 'Coral Rubble Colonized with Mixed Algae' was next (4.1% of the Lagoon), and occurred in a relatively continuous zone approximately 100 m inshore of the reef crest between Garapan and Sugar Dock. 'Seagrass (*E. acoroides*) on Sand' was the least abundant habitat mapped, comprising only 2.7% of the area. There were three main clusters of *E. acoroides*, all of which were nearshore off Tanapag, Red Beach, and near Smiling Cove Marina.

Prior to this assessment, the most recent habitat maps of the Saipan Lagoon were created based on 2001 IKONOS satellite imagery. Habitat changes since 2001 were evaluated by comparing the 2016 WV2 imagery and field data to the 2001 IKONOS image. These point-based comparisons revealed that 12% to 17% of locations experienced a habitat change and a measurable decline in seagrass cover. Fewest changes were observed on *E. acoroides* and pavement habitats (<3% of the observed changes). The greatest number of changes were among the *H. uninervis* and 'Mixed Algae and Seagrass on Sand' categories (20-30% of the observed changes). Based on these observations, change was examined in greater detail at five locations: north of the Sugar Dock, west of Red Beach, Garapan and Memorial Park, Tanapag, and around Mañagaha. In several cases, formerly dense seagrass meadows have been replaced with patchy sand and algal bottom. Expansion of several patch reefs comprised primarily of Staghorn *Acropora* was also evident, however, 2016 field images indicate that most of the Staghorn is now dead, covered with turf algae, and remaining live corals were bleached. Several locations experienced shoreline changes. Erosion of the beach around Memorial Park and accretion off the Hyatt Regency hotel and near Tanapag were evident between 2001 and 2016.

There are a wide range of applications for the 2016 habitat maps, satellite image, underwater videos, and other datasets. Habitat maps can help plan Lagoon uses to reduce conflict and minimize habitat impacts. They are also needed for measuring bottom features and estimating their economic value. Habitat maps are also ideal for planning spatial-management decisions on topics such as identifying and quantifying essential fish habitat, planning development to minimize habitats damage, monitoring habitat and shoreline changes, calculating damage and costs following ship grounding or other impacts, sample design for monitoring or scientific studies, and planning for marine managed areas.

The best way to use these highly detailed maps will be through GIS or other software that allows users to zoom in to a custom scale and manipulate the raster-based digital files. We also provide an on-line data viewer that allows users without any specialized software to access the habitat maps, the WV2 satellite image, the field data (including high definition underwater videos), as well as environmental predictor data. In addition, the WV2 image and composite habitat maps are provided in an atlas format at the end of this report. These products are archived at NOAA's National Centers for Environmental Information (NCEI) and are available through our website at: https://coastalscience.noaa.gov/projects/detail?key=271

Recreation on Managaha Island. Credit: M. Kendall and B. Costa

1.0 INTRODUCTION

*Staghon coral (*Acropora *species) in Saipan Lagoon. Credit: M. Kendall and B. Costa.*

The Lagoon along the western shore of Saipan, Commonwealth of the Northern Mariana islands (CNMI), encompasses a diverse coral ecosystem. The area plays a leading role in attracting nearly half a million tourists to the Island annually (Mariana Visitors Authority, 2017) and contributes to the economic value of Saipan's coral reefs, which was recently estimated at approximately $61 million per year (van Beukering et al., 2006). Habitats in the Lagoon are well protected from ocean swells by an emergent reef crest that extends along almost the entire 26 km of Saipan's western coastline. An approximately 4 km wide channel through the reef located mid-island is the only large opening to the open ocean. The reef crest is dominated by carbonate pavement, which absorbs most oceanic wave energy. Coral rubble is piled by wave action just behind the reef crest, followed by scattered patch reefs comprised of Staghorn *Acropora* thickets, *Isopora palifera*, massive *Porites* colonies, and other coral species. Bare sand, algal flats, and carbonate pavement dominate the floor of the central Lagoon. Seagrass meadows cover broad nearshore areas, often right up to the many sand beaches, where most development on the island is concentrated.

This diversity of habitats inside the Lagoon offers tourists abundant recreational opportunities, including snorkeling, diving, parasailing, kayaking, and use of personal watercraft (Northern Mariana Islands Tourism Master Plan, 2012). The Lagoon must also accommodate many other uses, such as commercial port operations at Tanapag Harbor and traditional fishing practices that are important to the local economy and culture (van Beukering et al., 2006). To manage this abundance of activity, the CNMI Government conducts periodic economic valuations (e.g., van Beukering et al., 2006) and has developed the Saipan Lagoon Use Management Plan (SLUMP, 2012). Understanding the present spatial distribution of important Lagoon habitats, and the potential human activities taking place within them, is needed to evaluate zoning scenarios, minimize user conflicts, ensure public safety, prevent environmental degradation, and estimate the economic value of this natural resource (van Beukering et al., 2006; CNMI and NOAA CRCP, 2010; SLUMP, 2012). The SLUMP and economic valuation studies are revised periodically (SLUMP: 1984, 1997, 2012, and underway in 2017, Economic Valuation: 2006 and slated for 2018) in response to changes in Lagoon habitats, user activities, and projections in tourist visitation. An accurate benthic habitat map inside the Lagoon is one of the primary pieces of information needed to update the 2017 SLUMP and the upcoming 2018 economic valuation.

Several maps of Saipan Lagoon have been produced since the 1940s. First, as part of a geological survey of Saipan, the Lagoon habitats and biota were mapped and characterized based on field data and oblique aerial photographs taken in 1949 (Figure 1.1; Cloud, 1959). The surveys resulted in a coarse depiction of reef zones, bottom types, and bathymetric contours that were the best maps available for over four decades. In 2005, the National Oceanic and Atmospheric Administration (NOAA) used a mosaic of IKONOS satellite imagery (4 meter raw multispectral and 1 meter pan sharpened pixel resolution) acquired in March 2001 to map the Lagoon as part of its comprehensive initiative to map all U.S. coral reef ecosystems (NOAA, 2002). This habitat map used a standard minimum mapping unit (MMU) of 1 acre (4,047 m^2), and a multi-layered classification scheme that included attributes for reef zone, geomorphological structure, and density of biological cover (Figure 1.2a; NOAA NCCOS, 2005). More recently, using the same 2001 IKONOS mosaic, Houk and van Woesik (2008) customized the classification scheme to describe the local features in Saipan Lagoon and to facilitate broad comparison to earlier maps (Cloud, 1959; Figure 1.2b).

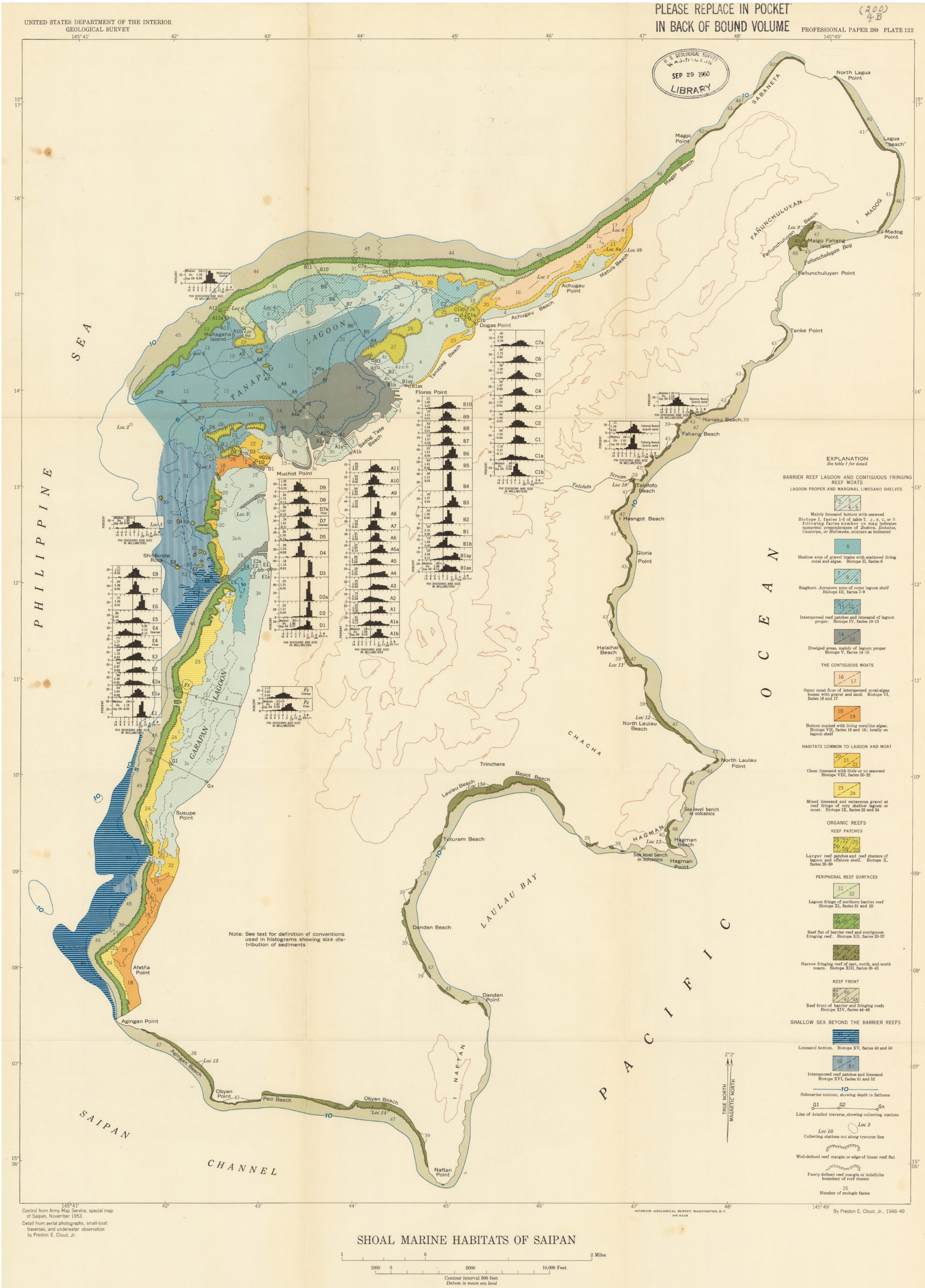


Figure 1.1. Lagoon features mapped from aerial photographs taken in 1949 from "Shoal Marine Habitats of Saipan", Plate 122. Source: Cloud, 1959.

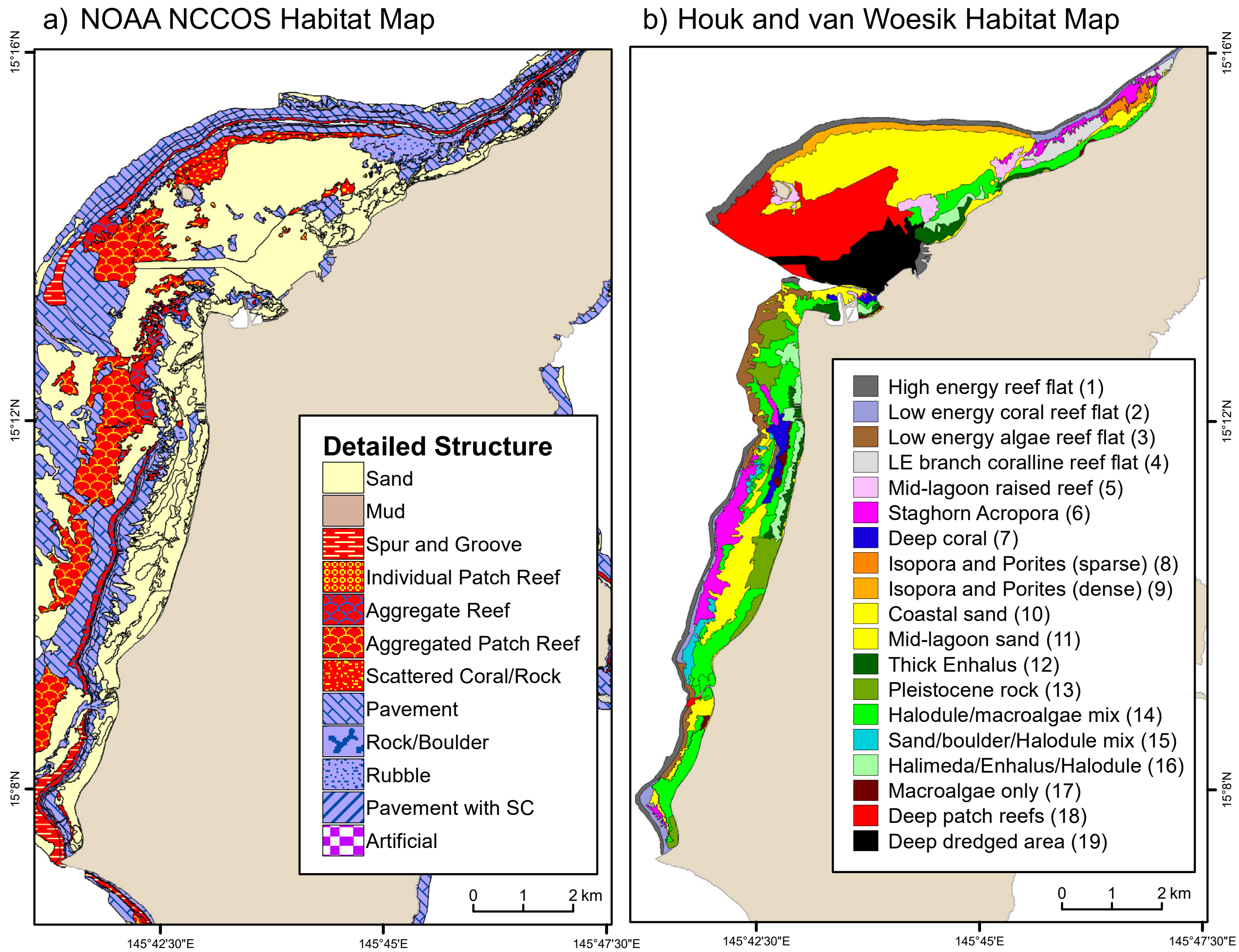


Figure 1.2. Lagoon features mapped from IKONOS satellite imagery taken in March 2001 by: a) NOAA NCCOS (NOAA NCCOS, 2005) and b) Houk and van Woesik (Houk and van Woesik, 2008).

Since the last mapping effort, several natural and anthropogenic processes have potentially altered Lagoon habitats. Perhaps most notably, powerful Category 4 Typhoon Soudelor directly struck Saipan in August 2015 and was the strongest storm to affect the island in decades (Guardand Lander, 2015). Storm induced waves, surge, and currents have the potential to re-arrange reef ecosystems through erosion and deposition of sand and rubble. Prior to Soudelor, recent studies have demonstrated how prevailing wind and wave climate, tidal currents, coastal modifications (e.g., channel dredging, debris removal, marina construction, and seagrass removal near hotels through CNMI Public Law 15-41) have interacted to influence erosion and deposition patterns around the hotel beaches (USACE 2004), at Memorial Park (Yuknavage and Palmer, 2010), and at Mañagaha Island (Fletcher et al., 2007). In addition, coral bleaching events due to elevated water temperatures have caused severe damage to broad areas in recent years. For example, a thermal-stress event in 2000 caused approximately 40% mortality in *Acropora* corals in the Saipan Lagoon (Houk and van Woesik, 2008). Consecutive years of bleaching in 2013-2014 reduced live coral to approximately 15% on those reefs (L. Johnston, unpublished monitoring data), and nearly all remaining *Acropora* corals were observed bleached in 2016 during the present study (M. Kendall, pers. obs.; B. Costa, pers. obs.). Drainage from urbanized watersheds has also been tied to changes in habitats over the last decade, such as shifts from seagrass to macroalgal dominated communities in the central Lagoon (Camacho, 2016). Recent storms and bleaching events, coupled with the ongoing impacts from coastal development and a rise in tourism, have increased the need for an updated map of Lagoon habitats.

Introduction

CNMI's top jurisdictional priority (as listed in NOAA Coral Reef Conservation Program's FY16 Internal Call for Proposals) requests: "support to update the habitat map for Saipan Lagoon, via remote sensing, ground truthing, and production of map layers, to support future monitoring and management efforts."

Scientist recording lagoon substrate and cover data off Wing Beach. Credit: M. Kendall and B. Costa.

The main objective of this project was to produce new, highly detailed maps of the extent and distribution of the bottom features within Saipan Lagoon (Figure 1.3). Recent advancements in not only the spatial and spectral resolution of satellite imagery, but also in computational power and complex mathematical models, enable a new generation of map products. The result is a dramatic increase in map detail, from approximately 4,047 m^2 polygons that were hand digitized and labelled in prior maps, to a 4 m^2 grid that was attributed using a more objective, repeatable analysis. The project was based on high resolution WorldView II (WV2) satellite imagery collected February 5, 2016 (post-Typhoon Soudelor). We used ground validation data from throughout the Lagoon to model the probability of occurrence of individual substrate and cover types, and to combine these model layers into a composite benthic map. The thematic and spatial accuracy of these products was evaluated using independent field data. Specifically, we sought to provide the following products:

1. Map of habitats in Saipan Lagoon;
2. Orthorectified, atmospheric- and water column-corrected WV2 satellite image;
3. Models predicting the occurrence of each individual bottom type throughout the Lagoon;
4. Field data and videos used to guide model development and measure accuracy of final maps ;
5. Estimated changes in the coverage and distribution of Lagoon habitats since 2001; and
6. A technical report (this document) describing the methods, results, and caveats for scientific and management applications of the products in Saipan Lagoon.

These products are freely available for viewing and are downloadable in GIS compatible formats from NOAA's National Centers for Coastal Ocean Science (https://coastalscience.noaa.gov/projects/detail?key=271).

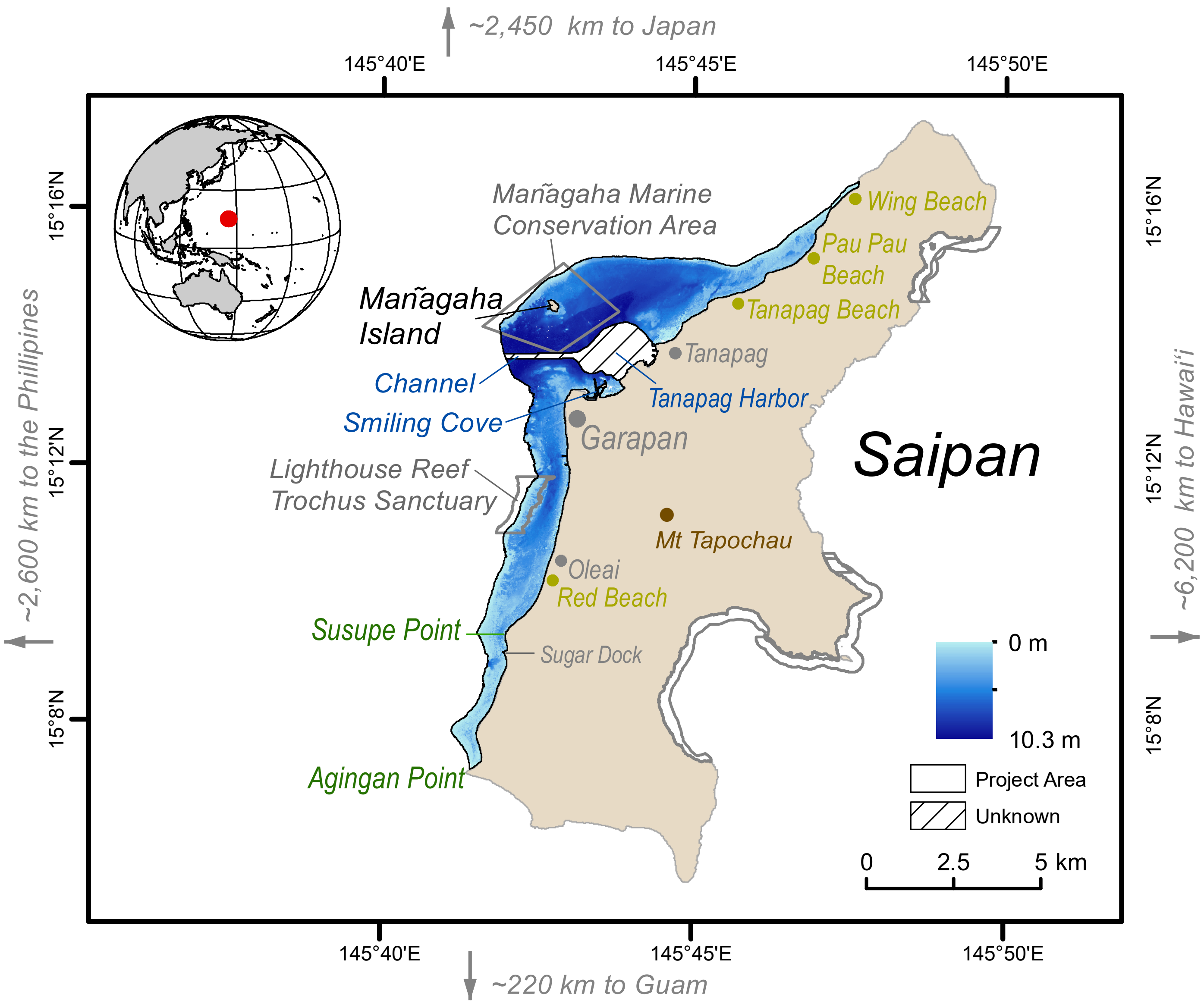


Figure 1.3. Key geographic features and place names.


Reef crest offshore at low tide. Credit: M. Kendall and B. Costa.

2.0 METHODS

Several processes were used to map and evaluate changes in habitats inside Saipan Lagoon. First, we describe the steps used during map development, including: customizing a habitat classification scheme; processing environmental variables, including satellite, topographic, and geographic predictors; collecting underwater video data; and creating habitat predictions using two modeling techniques called Boosted Regression Trees (BRTs) and Boosted Classification Trees (BCTs). Following that, we describe the independent dataset and process used to assess the performance and accuracy of the habitat models. Last, we explain the methods used to quantify changes in habitat distributions throughout the Lagoon and to highlight habitat shifts apparent between the 2001 and 2016 satellite images.

2.1 BENTHIC HABITAT CLASSIFICATION SCHEME

A habitat classification scheme allows scientists to systematically group bottom features based on their ecological characteristics. The classification scheme used here was developed by reviewing the previous habitat classifications applied in Saipan Lagoon (Cloud, 1959; NOAA NCCOS, 2005; Houk and van Woesik, 2008), consulting with local scientists and managers on their informational needs, and then devising and field-testing a draft scheme collaboratively with CNMI's Bureau of Environmental and Coastal Quality (BECQ) and NOAA's Pacific Island Regional Office (PIRO) in Saipan.

The scheme is based upon 12 benthic habitats, including five substrate (e.g., 'Sand', 'Pavement') and seven cover types (e.g., 'Live Coral (Staghorn *Acropora*)', 'Seagrass (*Enhalus acoroides*)') found in the Lagoon (Table 2.1). These 12 habitats guided collection of field data and development of two types of map products. First, the spatial distributions of these substrate and cover types were predicted inside the Lagoon using probability-of-occurrence models. This resulted in 12 map layers, where 2 x 2 m grid cells in the map denote the probability, from 0% to 100%, that a given habitat type is present. Second, the 12 probability-of-occurrence layers were used to create a composite (i.e., single layer) habitat map depicting commonly occurring combinations of substrates and cover types. Habitat classes in the composite map were defined based on cluster analysis of field data (performed in JMP v12 software). Clustering is a statistical technique used to identify groups of similar objects based on multiple attributes. Based on abundance data for substrate and cover at field sites, seven habitat types were defined (Table 2.2; e.g., 'Seagrass (*E. acoroides*) on Sand', and 'Coral Rubble Colonized with Mixed Algae'). These seven habitat classes, plus 'Unknown' and 'Artificial', were the basis of the composite map. This composite map was also translated into the Coastal and Marine Ecological Classification Standard (CMECS) to meet Federal Geographic Data Committee (FGDC) requirements (Tables 2.1-2.2; CMECS, 2017).

*Seagrass beds (*Enhalus acoroides*) in the southern Lagoon. Credit: M. Kendall and B. Costa.*

Methods

Table 2.1. Substrate and cover types used to describe individual benthic habitats. Equivalent Coastal and Marine Classification Standard (CMECS) classifications are suggested.

Habitats			Definition	CMECS IDs	CMECS Class
Substrate	1	Live Coral Reef (all spp.)	Presence of live coral reef. Comprised of live, upright hermatypic (reef-building) hard corals, including all hard coral species.	310, 421	Coral Reef Substrate (Substrate) and Mixed Shallow/Mesophotic Coral Reef Biota (Biotic)
Substrate	2	Upright Dead Coral Reef	Presence of dead hard coral reef that is still upright.	310, 501	Coral Reef Substrate (Substrate) and Colonized Shallow/Mesophotic Reef (Biotic)
Substrate	3	Pavement	Flat, low-relief or sloping solid carbonate rock with little or no fine-scale rugosity that is covered with algae, hard coral, gorgonians, zooanthids or other sessile vertebrates that are dense enough to partially obscure the underlying surface.	137, 310	Pavement Area (Geoform) and Coral Reef Substrate (Substrate)
Substrate	4	Coral Rubble	Dead, unstable coral rubble often colonized with filamentous or other macroalgae. This habitat often occurs landward of well developed reef formations.	311	Coral Rubble (Substrate)
Substrate	5	Sand	Coarse sediment typically found in areas exposed to currents or wave energy (Figure 2.19).	279	Sand (Substrate)
Cover	1	Live Coral (*I. palifera*)	Presence of live *Isopora palifera*	825	Branching *Acropora* reef (Biotic)
Cover	2	Live Coral (Staghorn *Acropora*)	Presence of live *Acropora aspera*, *A. formosa*, and *A. pulchra*	825	Branching *Acropora* reef (Biotic)
Cover	3	Live Coral (other spp.)	Presence of live scleractinian (hard) coral excluding *Acorpora* spp., *Isopora* spp., or *Porites* spp.	421	Mixed Shallow/Mesophotic Coral Reef Biota (Biotic)
Cover	4	Seagrass (*Enhalus acoroides*)	Presence of *Enhalus acoroides* seagrass	567	Seagrass Bed (Biotic)
Cover	5	Seagrass (*Halodule uninervis*)	Presence of *Halodule uninervis* seagrass	567	Seagrass Bed (Biotic)
Cover	6	Mixed Algae	Presence of mixed proportions of macroalgae (e.g., *Halimeda* spp.), turf algae, filamentous algae, crustose coralline algae and cyanobacteria.	433, 565, 560, 561, 567, 429	Benthic Macroalgae (Biotic), Turf Algal Bed (Biotic), Coralline/Crustose Algal Bed (Biotic), Filamentous Algal Bed (Biotic), Mat/Film Forming Microbes (Biotic)
Cover	7	Bare	No biological cover	No Equivalent	NULL
Cover	-	Unknown	Habitat unknown because water too deep or too turbid to see the seafloor	No Equivalent	NULL

Table 2.2. Commonly co-occurring substrate and cover types used to create a composite habitat map. Equivalent CMECS classifications are suggested.

Code	Habitat	Definition	CMECS ID (Geoform or Substrate)	CMECS Class (Geoform or Substrate)	CMECS ID (Biotic)	CMECS Class (Biotic)
1	Sand, Mixed Algae and Seagrass	>80% Sand covered with mixed proportions of *Halodule uninervis*, *Halimeda* spp., and other fleshy algae, remaining <20% is commonly rubble with turf algae	279	Sand	433, 565, 560, 561, 567, 429	Benthic Macroalgae, Turf Algal Bed, Coralline/Crustose Algal Bed, Filamentous Algal Bed, Seagrass Bed, Mat/Film Forming Microbes
2	Sand, Seagrass (*Halodule uninervis*)	>90% Sand that is >50% covered with *Halodule uninervis*	279	Sand	567	Seagrass Bed
3	Sand, Bare	>90% Sand that is >90% bare	279	Sand	NULL	NULL
4	Sand, Seagrass (*Enhalus acoroides*)	>90% Sand that is >90% covered with *Enhalus acoroides*	279	Sand	567	Seagrass Bed
5	Pavement, Mixed Algae	>50% Pavement covered with fleshy and turf algae, remaining <50% is mixed proportions of primarily rubble and sand	137, 310	Pavement Area (Geoform) and Coral Reef Substrate (Substrate)	433, 565, 560, 561	Benthic Macroalgae, Turf Algal Bed, Coralline/Crustose Algal Bed, Filamentous Algal Bed
6	Coral Rubble, Mixed Algae	>50% Coral rubble covered with turf and fleshy algae, remaining <50% is mixed proportions of primarily upright dead coral, pavement, and sand	311	Coral Rubble	433, 565, 560, 561	Benthic Macroalgae, Turf Algal Bed, Coralline/Crustose Algal Bed, Filamentous Algal Bed
7	Upright Dead and Live Coral Reef, Mixed Algae	>50% Live or upright dead coral covered with turf and fleshy algae, remaining <50% is primarily rubble	310	Coral Reef Substrate	501	Colonized Shallow/ Mesophotic Reef
8	Artificial	Man-made structures including shipwrecks, airplanes, tanks, breakwaters and piers	362	Anthropogenic Substrate	NULL	NULL
9	Unknown	Habitat unknown because water too deep or too turbid to see the seafloor	NULL	NULL	NULL	NULL

Methods

Predictor Data

Twenty-eight environmental variables were used to create the model predictions for individual substrate and cover types in the Lagoon. There were three broad categories of predictors: spectral bands from satellite imagery, geographic variables based on position in the Lagoon, and topographic surfaces based on depth and derivatives of depth.

Spectral Predictors

A WV2 satellite image acquired on February 5, 2016 at 10:44:34 am local time in Saipan was the basis of the mapping workflow (Figure 2.1). This excellent image was cloud free and had little to no sunglint, turbidity, or wave action. Twenty-four of the twenty-eight environmental predictors were derived from this satellite image. The WV2 sensor collects 8 multispectral bands (Table 2.3) at 2 x 2 m spatial resolutions, and a panchromatic (black and white) band at 0.5 x 0.5 m spatial resolution.

To correct geometric distortions caused by Saipan's mountainous topography, spectral bands were orthorectified using 20 ground control points (GCPs; NOAA NCCOS, 2002) and a digital elevation model (DEM; USACE, 2007; performed in ENVI 5.2: Orthorectify WorldView with Ground Control). The final orthorectified image (or orthoimage) was geo-referenced to the World Geodetic System 1984, Universal Transverse Mercator, Zone 55 North horizontal coordinate system (WGS84 UTM 55N). Positional accuracy was evaluated using an independent set of 25 GCPs collected using a Trimble GeoXH GPS receiver from August 3 to 8, 2016. GCPs were evenly distributed across the island and positioned on features that were clearly identifiable in the imagery, such as docks, parking lots, tennis courts, and other low-profile objects with distinct edges. Raw GPS data was post-processed and differentially corrected with Trimble Pathfinder Office software and data from the Mariana Island Continually Operating Reference System station. The final root mean square error (RMSE) for the WV2 image is 5.0 m.

Table 2.3. Wavelengths (nanometers) of eight multispectral bands collected by the WV2 sensor.

Band #	Band Name	Wavelength (nm)
1	CoastalBlue	400-450
2	Blue	450-510
3	Green	510-580
4	Yellow	585-625
5	Red	630-690
6	RedEdge	705-745
7	Near Infrared 1	770-895
8	Near Infrared 2	860-900

Satellite images capture the sunlight reflected by benthic habitats. This sunlight is partly absorbed and scattered as it passes through the atmosphere and the water column due to environmental factors such as aerosols or turbidity. These factors vary over space and time causing the same habitat to look different among locations. Such variability hinders the consistent and accurate discrimination of different habitat types (Mumby et al., 1998). To reduce these potential sources of error, the orthoimage was corrected for changing atmospheric and water column conditions (Lyzenga, 1978; Mumby and Edwards, 2000; performed using ENVI 5.2: THOR atmospheric correction tool and ArcGIS 10.4). These processes resulted in 15 atmospherically and water-column corrected bands or band-pairs for the mapping workflow (Figure 2.2).

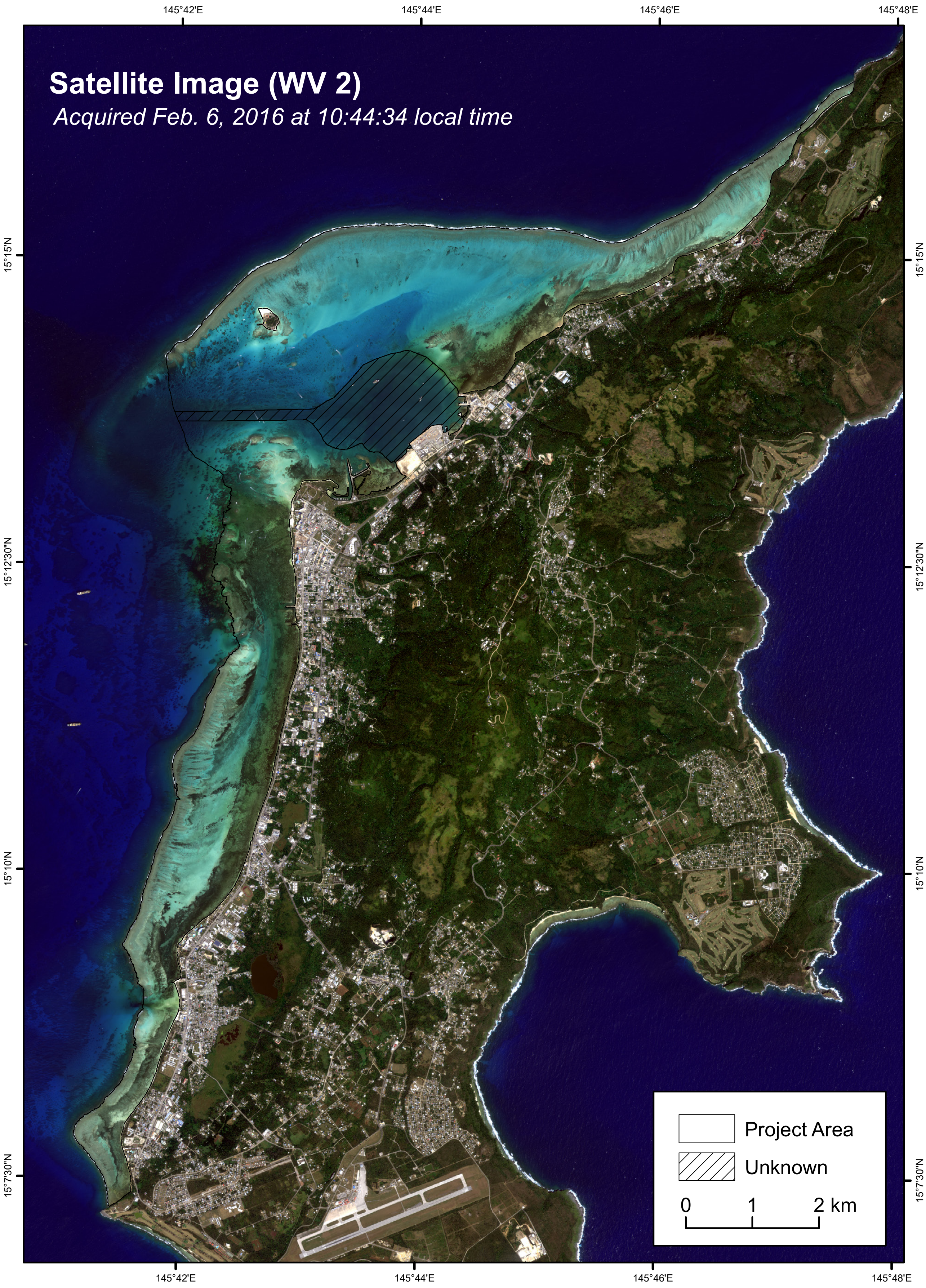


Figure 2.1. WV2 image acquired on February 5, 2016. Benthic habitats were characterized using this satellite image and several surfaces derived from it.

Methods

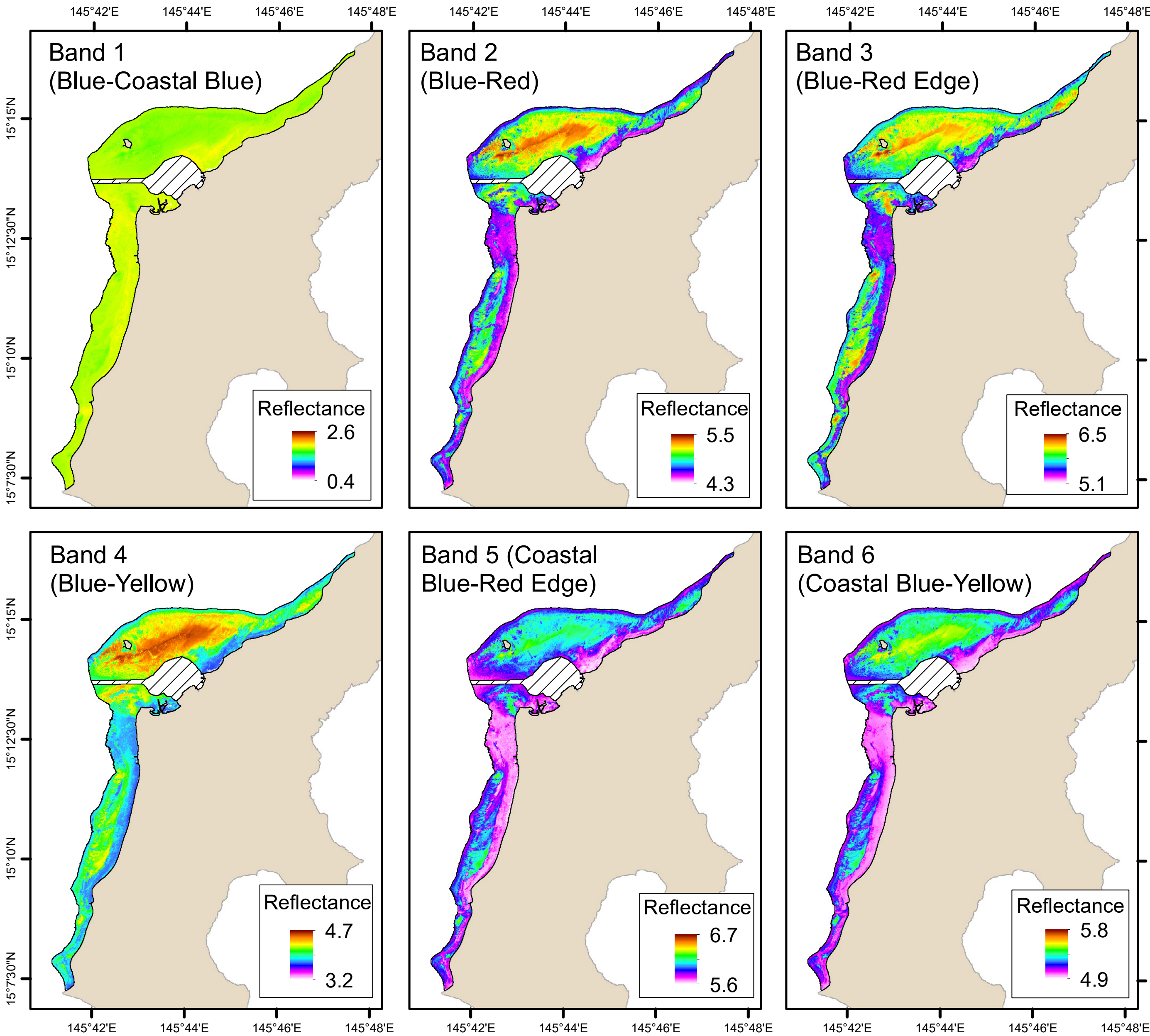


Figure 2.2. Maps depicting the orthorectified, atmospherically, and water-column corrected WV2 band pairs (1-6) used to create map products.

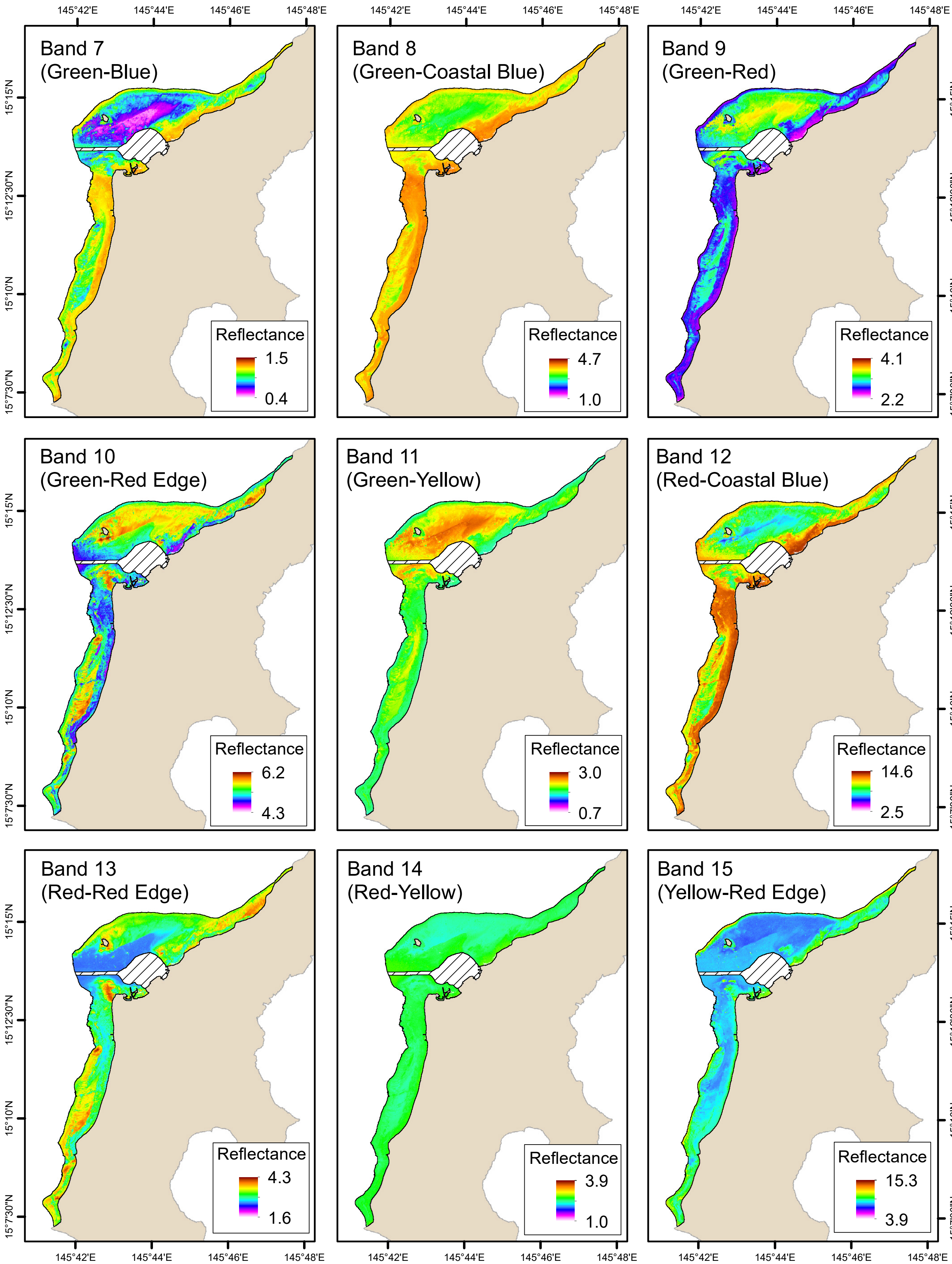


Figure 2.2. continued. Maps depicting the WV2 band pairs (7-15) used to create map products.

Methods

Geographic Predictors

Four geographic predictors were used to account for spatial variation in benthic habitats that was not explained by the spectral predictors. These included latitude (y), longitude (x), distance to shore, and distance to the reef crest (Figure 2.3; performed using ArcGIS 10.4's Euclidean Distance tool and Marine Geospatial Ecology Tools 0.8a64, MGET, 2016). The shoreline and reef crest locations were extracted from NOAA's previous benthic habitat map (NOAA NCCOS, 2005).

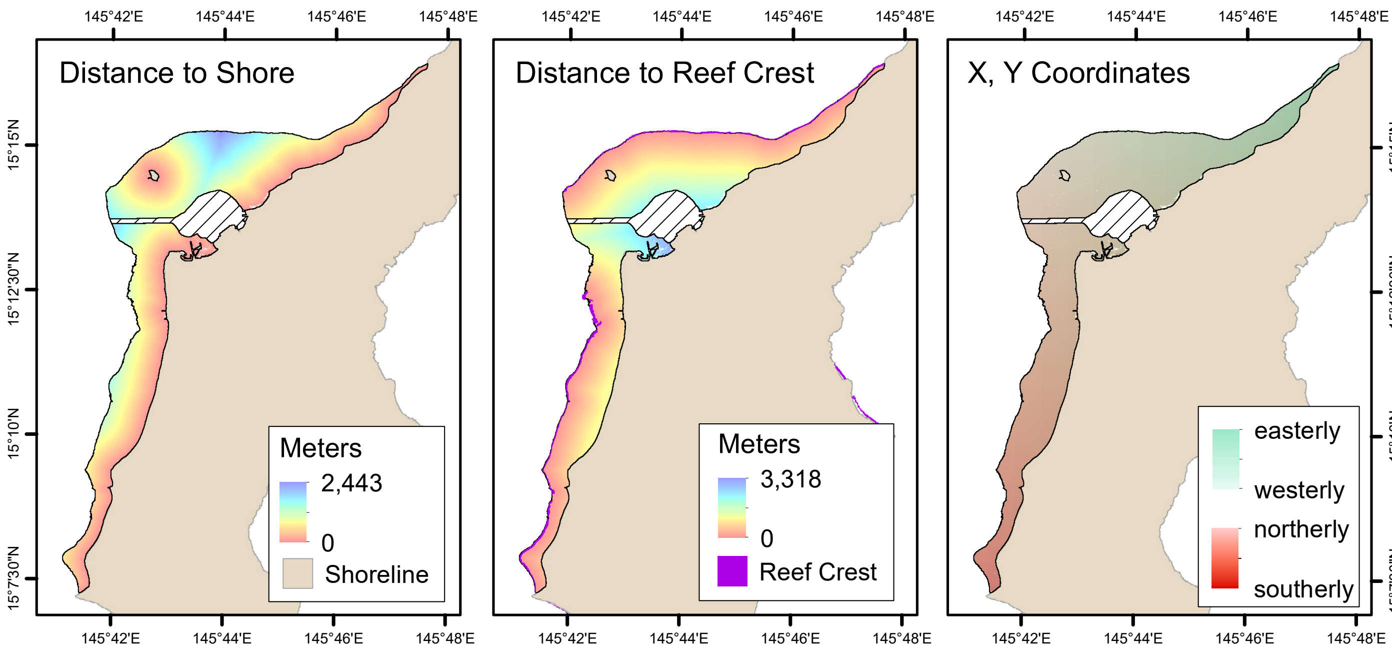


Figure 2.3. Maps depicting the geographic predictors used to create map products.

Topographic Predictors

Seafloor depth and topography are also known to be useful predictors of specific habitat types, such as sand, pavement, and reefs. We created a satellite-derived (SD) depth surface from the WV2 image using 243 depth measurements collected throughout the Lagoon using a graduated pole or weighted rope (see Section 2.3: Field Data). These depths were corrected for changes in tides (NOAA COOPS, 2016) and standardized to the Mean Lower Low Water (MLLW) tidal datum. Coordinates from the 243 sites were used to extract corresponding values from the four geographic and 15 spectral predictors.

Using these intersected data, we generated Boosted Regression Trees models (BRT; R Core Team, 2016) predicting MLLW depths (2 x 2 m grid) for the entire Lagoon (Figure 2.4). Precision (i.e., coefficient of variation or CV) associated with predicted depths was also calculated during the BRT modeling process (Figure 2.4). The accuracy of the SD depth surface was evaluated using an independent set of 273 *in situ* depth measurements (see Section 2.3: Field Data). This combination of recent high-resolution satellite imagery and post-Typhoon Soudelor field data offered an improvement over a previously available SD depth surface (Bindel et al., 2013).

Eight topographic surfaces describing the complexity of the seafloor were derived from the SD depth surface: (1) Depth (Coefficient of Variation), (2) Depth (Standard Deviation), (3) Total Curvature, (4) Plan Curvature, (5) Profile Curvature, (6) Rugosity, (7) Slope, and (8) Slope Rate of Change (Figure 2.4). Each topographic surface was calculated using a square 3 x 3 cell neighborhood, where the central pixel in the neighborhood was assigned the calculated value (performed using ArcGIS 10.4: DEM surface tools, see Jenness, 2016).

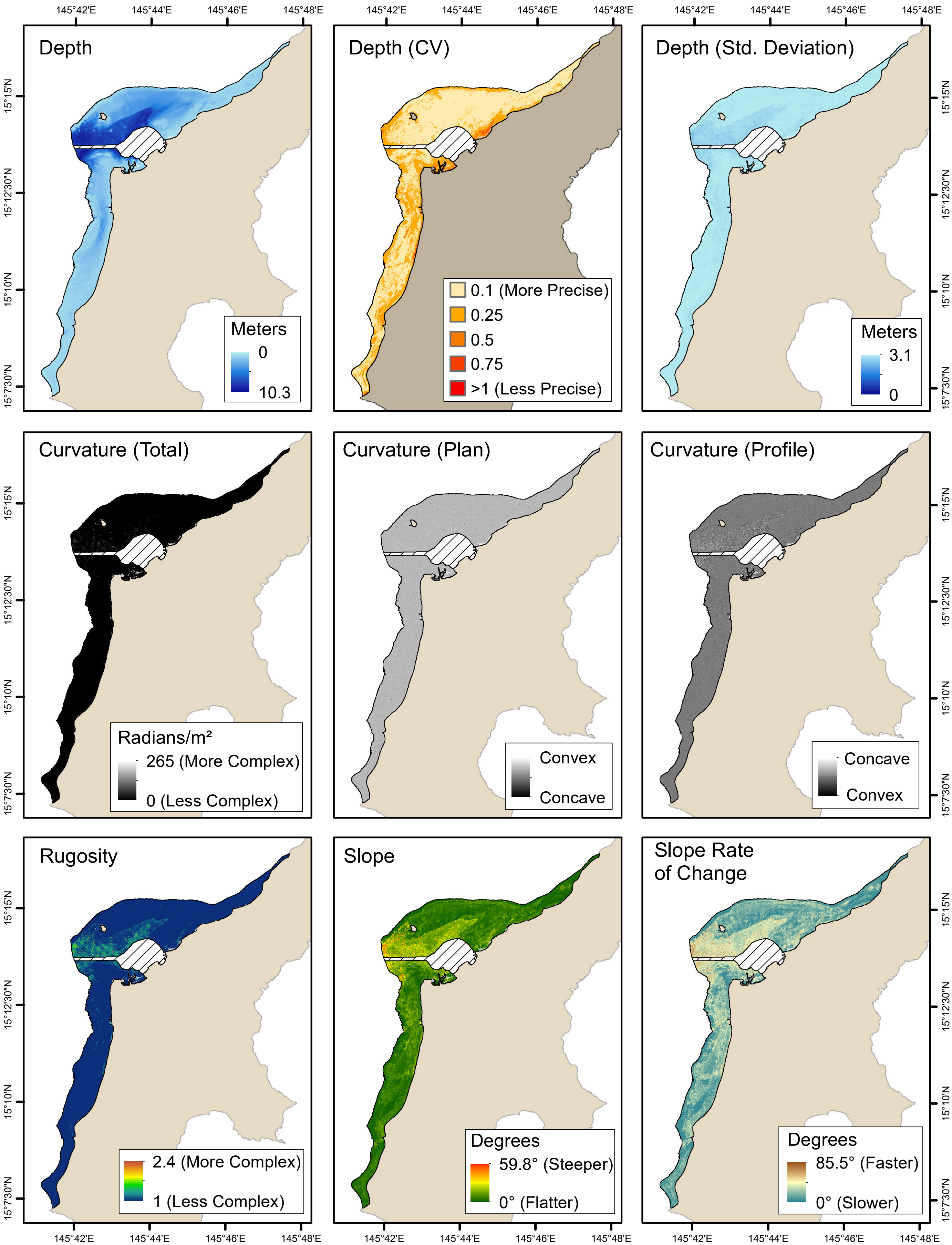


Figure 2.4. Maps derived from the SD depth surface depicting the topography of the seafloor and used to create map products.

Methods

2.3 FIELD DATA

Underwater video and photographs were acquired between 28 July 2016 and 9 August 2016 at nearly 600 sites throughout the Lagoon to document the presence and percent-cover of substrate and cover types in the classification scheme (Table 2.1). Two sets of independent data were collected, one for Ground Validation (n = 292) and the other for Accuracy Assessment (n = 273).

Ground Validation (GV) Training Dataset

GV data are the basis for correlating observed substrate and cover types with their associated values in the predictor datasets. Ultimately, the data are used to train and optimize mathematical models that predict habitats throughout the Lagoon. Locations of the GV sites were selected deliberately to include the full range of habitats, depths, and environmental settings found in the Lagoon. We collected underwater videos and photos at 243 GV sites (Figure 2.5) using the process described below. This was supplemented by an additional 49 GV sites collected in July-August 2016 in collaboration with staff from NOAA PIRO and CRCP in Saipan.

Accuracy Assessment (AA) Test Dataset

AA data are used to independently evaluate the performance and accuracy of the predictive models and the composite habitat map. AA sites were randomly positioned and stratified using a preliminary map comprised of 9 habitat strata. These habitat strata were mapped using previously existing GV data from earlier mapping projects (NOAA NCCOS, 2005; Houk and van Woesik, 2008). A set of 350 sites were identified by randomly distributing 35 points within each of the 9 habitat strata and then adding an additional 35 points within the 'Sand' stratum due to its extent throughout the Lagoon (performed using ArcGIS 10.4: Create Accuracy Assessment Points tool; Congalton and Green, 1999). Due to inclement weather and visibility in the dredged area of Tanapag Harbor, underwater videos and habitat data were collected at 273 of the 350 AA sites in the field (Figure 2.5).

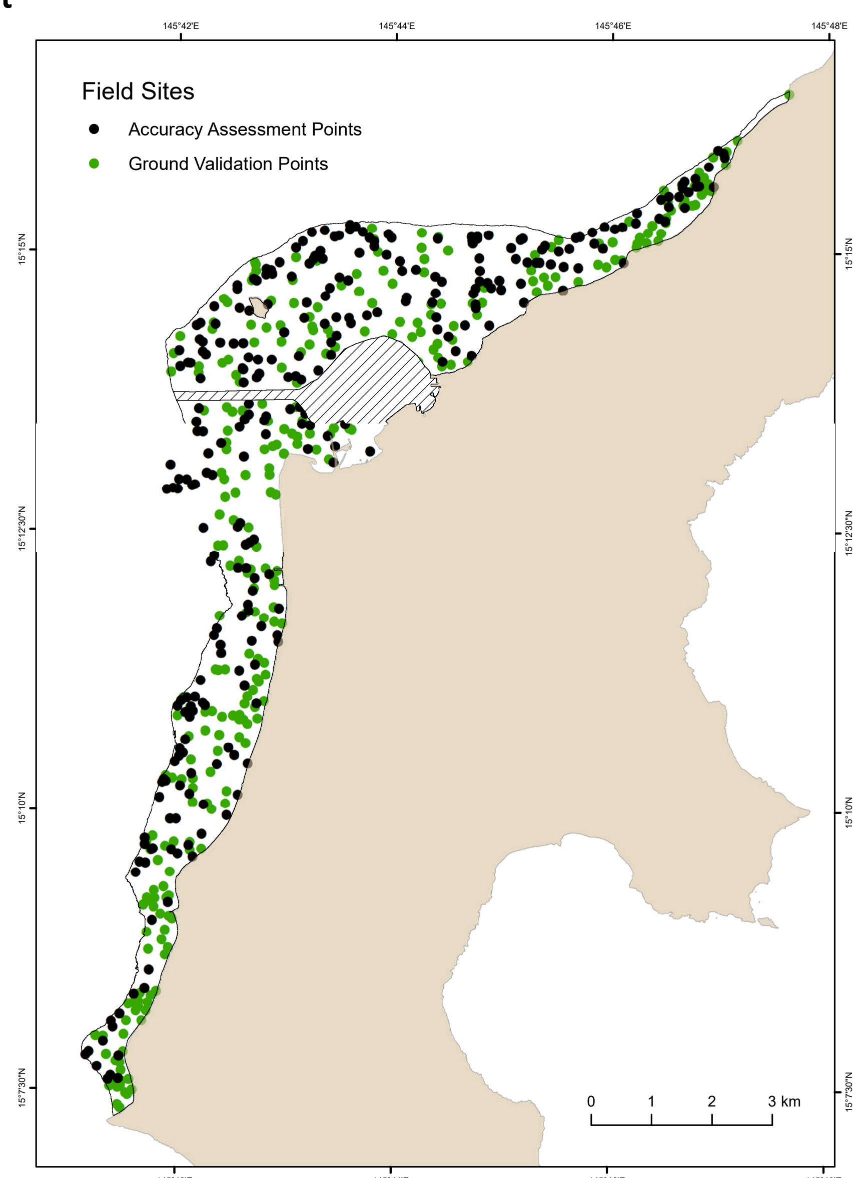


Figure 2.5. Locations of GV and AA underwater videos, photographs, and field data collection.

The process for collecting both the GV and AA data was identical at each field site. Sites were typically accessed via small boat, kayak, or wading. A hand-held Garmin 76 global positioning system (GPS) unit was used for navigation to site coordinates. At each site, two GoPro HERO4 Black cameras were deployed

on an aluminum pole for sites <9 m deep (Figure 2.6a), and on a rope with a rotating camera system (RCS) for deeper sites (Figure 2.6b). On both deployment systems, a downward facing camera was fixed at 1 m above the bottom to standardize the field of view to encompass approximately 1 m^2 of seafloor, and an oblique facing camera captured surrounding habitats. Both cameras collected high definition (HD) video (30 frames per second, 1,920 x 1,080) and photographs (7 megapixels, 3,008 x 1,692 pixels) every 5 seconds for approximately 1 minute. Supplementary photographs were also taken using a hand-held digital camera at many sites. Once the cameras were deployed, our precise location was recorded every five seconds for the duration of the video using a Trimble GeoXH GPS receiver (Figure 2.6c). Substrate and cover types from the classification scheme were estimated to the nearest 10% in real time using a look bucket (Figure 2.6d), or free diving, and then recorded on a paper datasheet. Once back in the office, GPS data were post-processed and differentially corrected using the Mariana Island Continually Operating Reference System station. Average positions were calculated for each site and then visually checked for correspondence between positions of video and satellite imagery. All field videos were reviewed for quality control of image clarity and classification values. For input and assessment of predictive models, abundances for the five substrate and seven cover types were converted to presences (1) and absences (0). Multiple substrate and cover types could be present at each site.

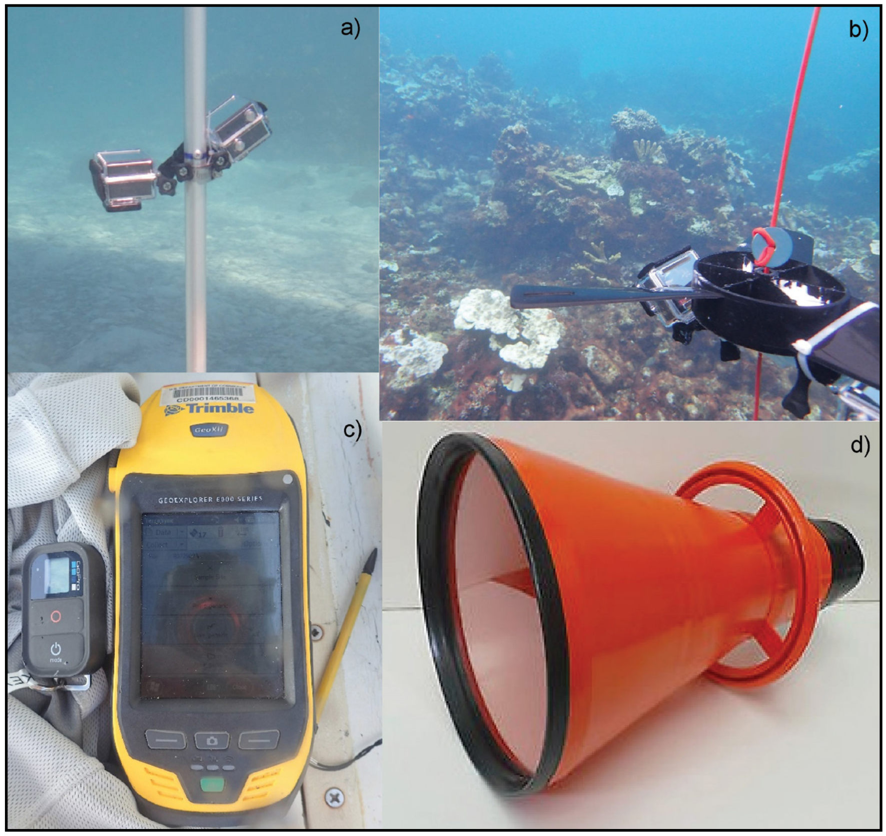


Figure 2.6. Equipment used to collect GV and AA data in the field: a) Go Pro HERO4 Black cameras mounted on aluminum pole; b) GoPro HERO4 Black cameras on rotating/rope mount; c) Trimble GPS unit and GoPro remote; and d) look bucket.

2.4 PREDICTING AND CLASSIFYING BENTHIC HABITATS

Boosted regression trees (BRTs) and boosted classification trees (BCTs) are machine-learning techniques that combine regression or classification trees with boosting to model the complex, non-linear relationships between bottom types and environmental variables (See Glossary for model terminology). BRTs and BCTs model these complex relationships by developing many (hundreds to thousands) simple classification or regression (tree) models. Classification and regression trees (Breiman et al., 1984) relate a response to predictors by iteratively splitting the data into two homogenous groups. These models are built in a stage-wise fashion, where existing trees are left unchanged and the variance remaining from the last tree is used to fit the next one. This stage-wise process is called boosting. A random subset of data is used to fit a model at each stage. This randomization helps improve model performance (Friedman, 2002; Elith et al., 2008). These simple models are then combined linearly to produce one final combined model (Elith et al., 2008). The fitted values in this combined model are more stable than values from an individual model, improving its overall predictive performance (Friedman, 2002; Elith et al., 2006, Elith et al., 2008).

Methods

We developed 12 separate, spatial predictions for the presence of substrate and cover types using BRTs and one composite habitat map using BCTs. We used these modeling techniques because they can deal with data that is not normally distributed (Elith et al., 2008), and are robust to missing data values (Breiman et al., 1984; Elith et al., 2008). They can also handle many types of response variables (presence/absence, count, diversity, and abundance), environmental predictors (numeric, binary, or categorical), and interactions among predictors (De'ath, 2007; Elith et al., 2008). These techniques also compare favorably to other modeling techniques both in predictive performance and accuracy (De'ath and Fabricuis, 2000; Elith et al., 2006; Elith et al., 2008).

Three main steps were used to create habitat predictions and a composite habitat map for Saipan Lagoon: (1) preparing the data, (2) creating and evaluating substrate and cover models and spatial predictions, and (3) creating and evaluating a composite habitat map (Figure 2.7). This work was conducted primarily in ArcGIS 10.4 (ESRI, 2016) and R software (R Core Team, 2016) using the dismo (Hijmans et al., 2014), caret (Kuhn et al., 2016), and raster (Hijmans, 2014) packages.

Step 1 – Prepare Input Data

We used the presence and absence of five substrate and seven cover types as the response variable in the BRT modeling process. This binary response variable was modeled using a binomial (two groups) distribution. No transformations were applied. All of the 28 environmental (i.e., spectral, geographic, and topographic) predictors were numeric. We conducted pairwise testing to identify and remove predictors that were highly correlated (i.e., Spearman rank $r \geq 0.9$ or $r \leq -0.9$) with three or more other predictors. After reducing the number of predictors from 28 to 24, we intersected the GV sites (i.e., locations denoting the presence and absence of substrate and cover types) with the remaining predictors to extract their value at each location. This spatial intersection combined the GV and predictor datasets into a single table.

Step 2 – Create & Evaluate Models and Spatial Predictions

In this step, we fit and optimized the BRT model parameters in R (R Core Team, 2016) using the dismo package (Hijmans et al., 2014). Several model parameters were tested during this process, including the learning rate (*lr*), tree complexity (*tc*), and bag fraction (*bf*). Learning rate (*lr*) controls how much each tree contributes to the model. The larger the learning rate, the more each tree contributes to the model. Tree complexity (*tc*) dictates how many nodes (splits) there are in a tree. The more splits there are, the more complex the model. Bag fraction (*bf*) specifies the proportion of data that is randomly chosen at each step. The larger the bag fraction, the more data that is available to train the model at each step.

For each of the five substrate and seven cover types, we tested 36 combinations of *lr*, *tc*, and *bf* (Table 2.4). We used k-fold cross validation (kCV) to identify the combinations of *lr*, *tc*, and *bf* that created the model with the smallest amount of error. Here, the kCV process divided the input table into 10 folds (i.e., 10 data subsets). Nine of these were used to create models, while the one remaining was used to evaluate the model's performance.

Table 2.4. Suite of boosted regression tree (BRT) model parameters and values tested.

Regularization Parameters	Parameters Tested	Definition	Impact	Example
Learning Rate (*lr*)	0.01, 0.001, 0.005	Determines contribution of each tree to the growing model	Decreasing (slowing) *lr* increases the number of trees required for optimal prediction	*lr* = 0.005 will grow more trees than *lr* = 0.01
Tree Complexity (*tc*)	2, 3, 4, 5, 10, 20	Controls how many predictor interactions are fitted in a tree	Decreasing *tc* will shrink the size (number of nodes) in a tree	*tc* = 20 will grow larger trees (with more nodes) than *tc* = 2
Bag Fraction (*bf*)	0.5, 0.75	Controls proportion of data randomly selected to build each tree	Decreasing *bf* will reduce the number of points randomly used to build a tree	*bf* = 0.5 will randomly sample 25% fewer data points than *bf* = 0.75

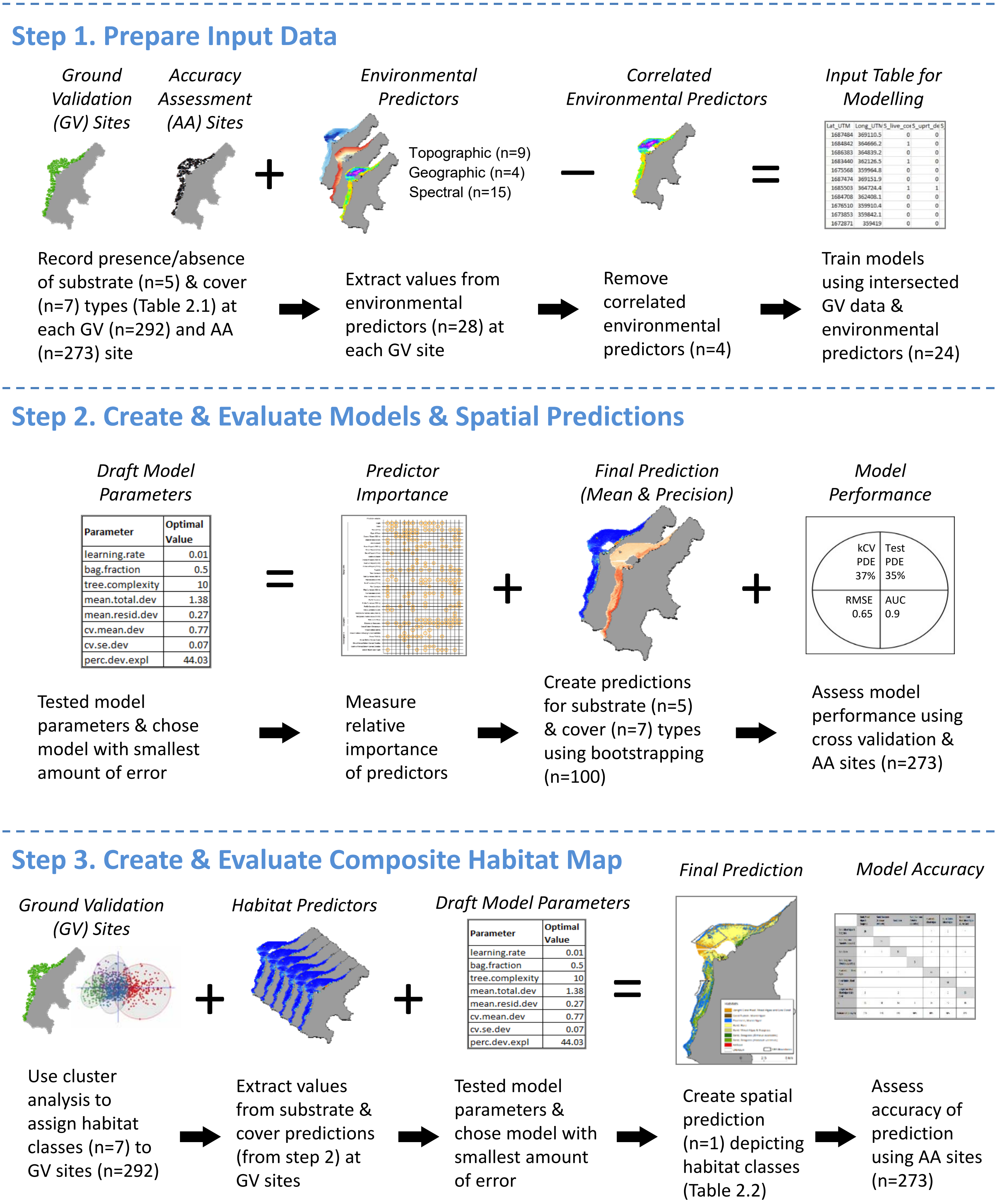


Figure 2.7. Diagram depicting steps in modeling process to predict substrate and cover distributions and develop a composite habitat map.

Methods

This process was repeated 10 times (i.e., one time for each fold) x 36 model parameter combinations x 12 substrate and cover types (n = 4,320 models total). We measured model performance using the percent deviance explained (PDE) averaged across the 10 folds. PDE is the amount (%) of variation explained in the response data with the average trend removed. PDE values normally range between 0 and 100%, with higher values indicating better model performance and lower error. We selected the 12 models (i.e., one model for each substrate and cover type) with the highest PDE. The remaining models were discarded.

Next, we used the 12 best models to summarize the relative importance of each environmental predictor for each substrate and cover type. Relative importance describes how often a predictor is used for tree splitting (Elith et al., 2008), and can identify potential ecological drivers that influence the distribution of habitats. These 12 best models were also used to predict the distribution of the five substrate and seven cover types throughout the Lagoon (performed using the raster package in R; Hijmans, 2014; R Core Team, 2016). These predictions describe the probability-of-occurrence for each habitat (i.e., the likelihood [%] that a particular substrate or cover type is present). Larger probabilities indicate it is more likely the habitat is present. We also quantified the precision associated with each probability of occurrence prediction, reported as the coefficient of variation (CV). CV is a measure of model precision, and represents the standard deviation as a proportion of the mean (Leathwick et al., 2006). Larger CVs indicate there is more uncertainty associated with the spatial prediction.

These probabilities and precisions represent the average of, and variation in, 100 model iterations created for each substrate and cover type using bootstrapping. Bootstrapping is a data resampling technique that randomly selects a subset of response data, fits a model to that data subset, and then applies that model to create a spatial prediction. The data subset is put back into the whole dataset at the end of the process, and a new random subset is selected. This process occurs iteratively to better understand how the models and spatial predictions change when given different response datasets, and whether specific response data subsets bias the model results.

Last in this step, we assessed the performance of the substrate and cover type predictions using four different metrics: (1) kCV PDE, (2) test PDE, (3) RMSE, and (4) Receiver Operating Characteristic (ROC) Area Under the Curve (AUC). kCV PDE was calculated during k-fold cross validation by comparing the observed values (in one randomly chosen validation fold) to the predicted values (from the models developed using the remaining nine training folds). Test PDE, AUC, and RMSE were independently calculated using the independent AA dataset (Section 2.3: Field Data). Test PDE, like kCV PDE, is the amount (%) of variation explained in the response data with the average trend removed. PDE values normally range between 0 and 100%, with higher values indicating better model performance. Conversely, RMSE measures the error associated with a model by calculating the difference between the predicted values (extracted from the model) and the observed values (extracted from the underwater videos). Lower RMSE denotes less error.

ROC curves measure a model's predictive performance in a different way compared to PDE and RMSE. Specifically, ROC curves compare a model's sensitivity (i.e., true positive prediction rate) to its specificity (i.e., false positive prediction rate). This rate depends on the choice of a particular probability of occurrence threshold above which substrate or cover types are classified as "present" and below which they are classified as "absent." Area Under the ROC Curve (AUC) does not require selecting a threshold, and can be used to measure the overall predictive performance of a model (compared to a random guess). AUC values ranging from 0.7 to 0.8 denote "good" model performance; values from 0.8 to 0.9 denote "excellent" model performance, and values greater than 0.9 denote "outstanding" model performance (Hosmer and Lemeshow, 2000). AUC values at or below 0.5 indicate that the model's prediction was no better than one created by chance alone.

We calculated these four different metrics because they describe model performance in different ways, and when viewed together, can provide a more thorough understanding of the models limitations. For example, models with higher kCV PDE, test PDE, and AUC values, but lower RMSE suggest they can be used with greater confidence because they correctly explain more variation in the response data with lower amounts of error.

Step 3 – Create and Evaluate Habitat Map

In this step, BCTs were used to develop a composite habitat map depicting the distribution of the seven habitats identified by cluster analysis (Section 2.1 Benthic Habitat Classification Scheme, Table 2.2). We modeled this response variable using a multinomial (many groups) distribution. The 12 probability of occurrence maps for individual substrate and cover types were used as predictors. No predictors were eliminated prior to modeling since they were not highly correlated (Spearman rank $R < 0.9$ or $R > -0.9$). We intersected the 292 GV sites (each of which were assigned one of the seven habitat types) with the 12 probability of occurrence predictions to extract their value at each location. This spatial intersection combined the GV and predictor values into a single table.

Next, model parameters for BCTs were fit and optimized in R (R Core Team, 2016) using the caret package (Kuhn et al., 2016). We tested 180 combinations of *lr*, *tc*, number of trees (*n.trees*), and minimum terminal node size (*n.minobsinnode*; Table 2.5). The *lr* and *tc* parameters are the same as those used to develop BRTs above. Number of trees denotes the number of classification trees that are fitted to the response data. The minimum terminal node size tells the modelling process when to stop splitting the response data, and denotes the number of observations (e.g., 3, 5, or 10) for each end point in a classification. We used k-fold cross validation (kCV) to identify the combinations of *lr*, *tc*, *n.trees* and *n.minobsinnode* that created the model with the smallest amount of error. We calculated kCV PDE to identify the parameter combination that created the highest performing model. This highest performing model was then applied spatially to create the composite habitat map throughout the Lagoon.

Minor manual editing was needed to improve this composite map. These edits included filling in small gaps along the shoreline that were removed during processing of the satellite image, editing obvious defects in predictions, removing boats and their wakes, and adding two additional map classes. These additional map classes described 'Unknown' areas (i.e., the deep, dredged basin and channel into the harbor) and 'Artificial' features (i.e., shipwrecks, airplanes, tanks, and piers; Table 2.2). Manual edits consisted of approximately 4% of the total mapped area.

Table 2.5. Suite of boosted classification tree (BCT) model parameters and values tested.

Regularization Parameters	Parameters Tested	Definition	Impact	Example
Learning Rate (*lr*)	0.01, 0.001, 0.005	Determines contribution of each tree to the growing model	Decreasing (slowing) *lr* increases the number of trees required for optimal prediction	*lr* = 0.005 will grow more trees than *lr* = 0.01
Tree Complexity (*tc*)	2, 5, 10	Controls how many predictor interactions are fitted in a tree	Decreasing *tc* will shrink the size (number of nodes) in a tree	*tc* = 20 will grow larger trees (with more nodes) than *tc* = 2
Number of Trees (*n.trees*)	500, 750, 1000, 2000, 3000	Describes the number of classification trees that are fitted to the response data	More classification trees will create more complex models (at the risk of overfitting the data)	*n.trees* = 500 will grow 500 classification trees
Minimum Terminal Node Size (*n.minobsinnode*)	3, 5, 10	Describes the number of observations at each endpoint in a classification tree	A lower number of observations will increase the risk of overfitting the model	*n.minobsinnode* = 3 will stop fitting when a classification tree has 3 observations

2.5 MEASURING THEMATIC ACCURACY

The thematic accuracy of the composite habitat map was assessed using the 273 AA sites. We grouped AA sites into the same seven habitats identified by the cluster analysis. Sites were considered correct if the same habitat was present within four meters (two pixels) of the AA site due to the positional accuracy of the WV2 orthoimage. A matrix was developed from the 273 AA points describing the composite maps' overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA; Story and Congalton, 1986). This matrix was constructed as a square array of numbers arranged in rows (map classification) and columns (accuracy assessment classification). The OA was calculated as the sum of the major diagonal (i.e., correct classifications), divided by the total number of AA samples.

The PA and UA were calculated to describe the thematic accuracy of individual map categories. PA describes errors due to omission, and is a measure of how well the cartographer classified a particular habitat (e.g., the percent of times that a site recorded as sand in the field was correctly mapped as sand). UA describes commission errors, and is a measure of how often certain habitat types were classified correctly (e.g., the percentage of times that a pixel classified as sand was actually verified as sand in the field). Each diagonal element was divided by the column total (n_i) to yield a producer's accuracy, and by the row total (n_j) to yield a user's accuracy. We also calculated the Tau coefficient to account for the random, chance agreement between the map and AA data (Ma and Redmond, 1995). The probability of random agreement decreases as the number of habitat classes increases.

While stratification helps ensure all habitat classes are adequately evaluated, it has the undesired effect of introducing bias into the confusion matrix. This bias is due to different sizes (km^2) of areas occupied by each habitat class (Card, 1982), causing rare habitats (e.g., live coral) to be sampled at a greater density than common habitats (e.g., sand). This sampling bias was removed using the method of Card (1982), which uses the proportion (%) of the map occupied by each habitat to correct thematic accuracies. These proportions were also used to compute confidence intervals for the overall accuracy (Card, 1982; Congalton and Green, 1999). For more information about these calculations and the equations, please see Kågesten et al. (2015) or Costa et al. (2013). As a last step, any errors identified during the accuracy assessment were also corrected in the final composite map.

2.6 MEASURING CHANGES IN LAGOON HABITATS 2001-2016

The scientists and managers working in the Lagoon need a better understanding of how habitats have changed since the previous maps were created (NOAA NCCOS, 2005; Houk and van Woesik, 2008). Unfortunately, it is not straightforward to simply compare maps that were created based on the March 2001 IKONOS mosaic to those created from the February 2016 WV2 image. These maps differ in geopositioning, resolution, and quality of source imagery, minimum mapping unit, classification scheme, and also the distribution and quantity of GV data used to create them. Similar habitat classes could be used to quantitatively compare such disparate maps, but any differences will be due to both actual changes in habitats and differences in the methods used to create the maps (Kendall and Miller, 2008; Houk and van Woesik; 2008).

To avoid these issues, habitat changes in Lagoon features from the 2001 IKONOS mosaic (NOAA NCCOS, 2001) to the 2016 WV2 image were based on image to image comparisons. First, the IKONOS mosaic was rubber-sheeted (i.e., coregistered) to match the more precise georeferencing of the WV2 scene using the 2016 GCPs (performed in ENVI Classic 5.2: Warp from GCPs: Image to Image; Figure 2.8). The final RMSE was 3.8 m for the 2001 IKONOS mosaic, minimizing the chance that the spatial positioning of the satellite images would influence the results.

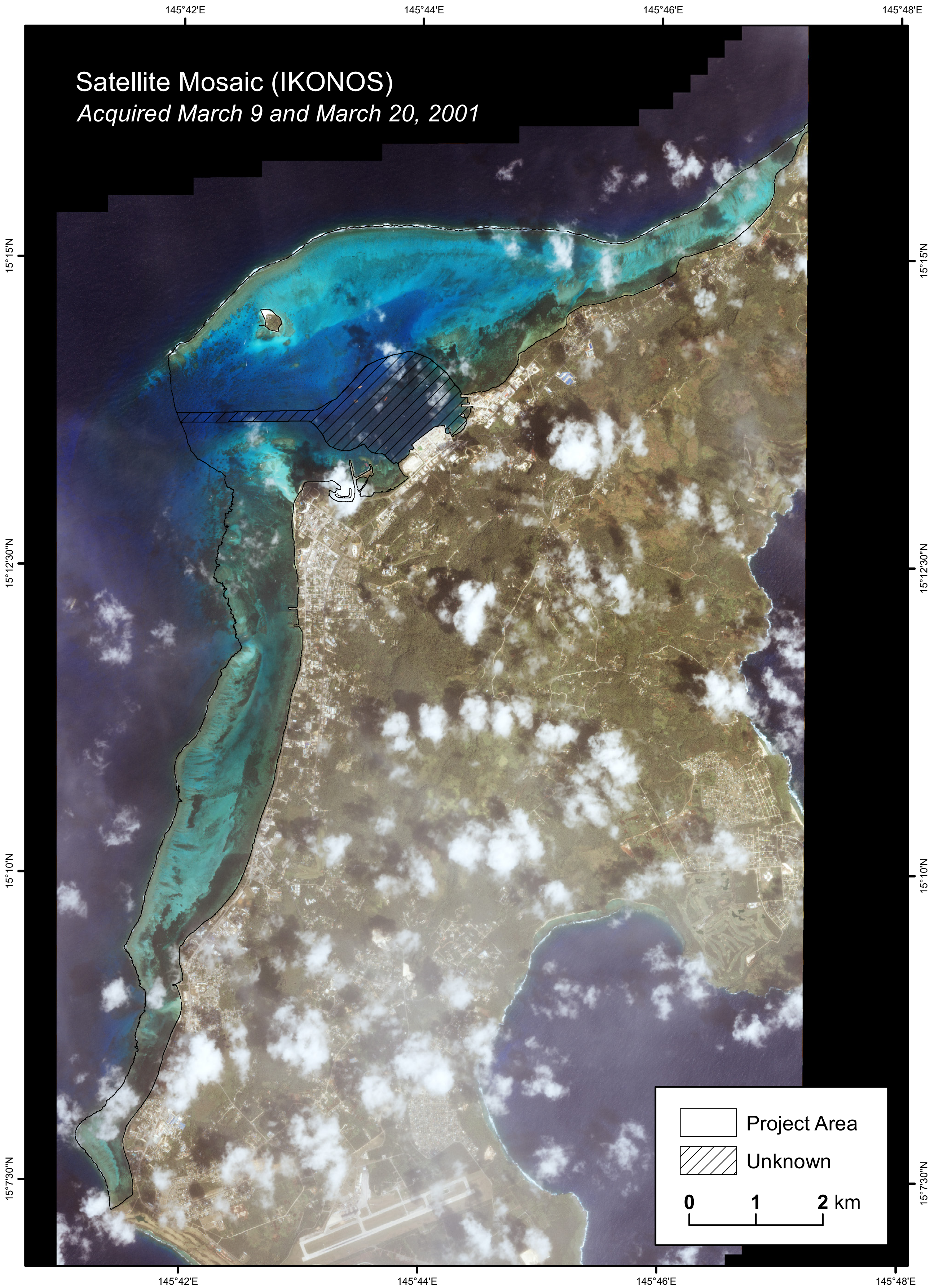


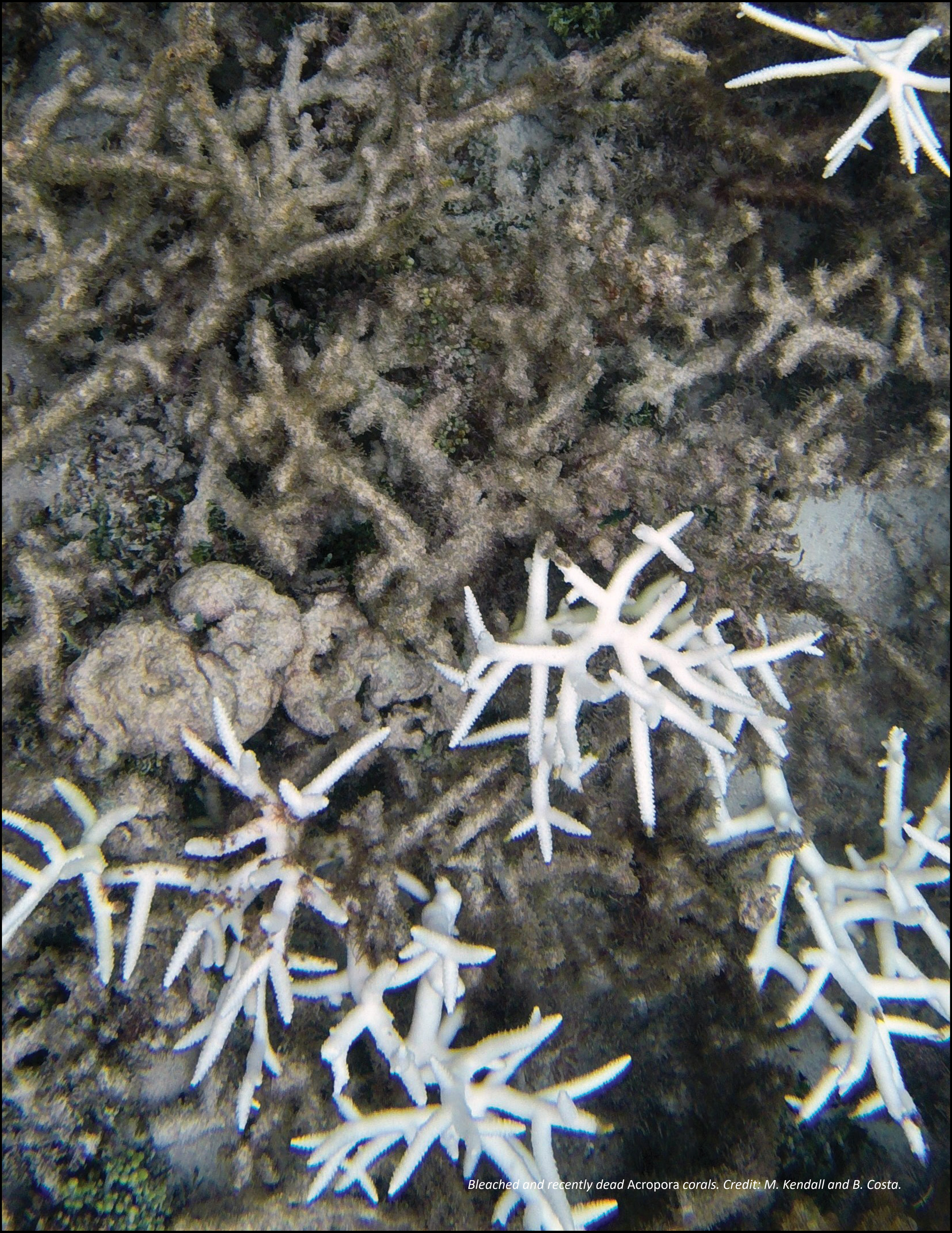
Figure 2.8. IKONOS mosaic created from satellite images collected on March 9 and March 20, 2001. We compared this satellite mosaic to the 2016 WV2 satellite image to quantify changes in benthic habitats over time.

Methods

Three techniques were used to evaluate various aspects of habitat change in the Lagoon. First, the AA data collected at field sites in 2016 (n = 273) was overlayed on the 2001 IKONOS mosaic. Specifically, the seven habitat classes used in the composite habitat map were compared to the underlying features in the IKONOS mosaic. Habitats from the IKONOS image were visually classified into each of the seven habitat classes at a scale of 1:1000, and scored in a matrix of matches and mis-matches. Locations with cloud interference in the 2001 IKONOS were omitted from analysis (reduced to n = 252). A conservative approach was taken wherein only clear, unambiguous changes from one habitat type to another were scored as mis-matches. This approach is robust and maximizes the scope of inference throughout the Lagoon, because the 2016 field data represents highly confident ground truth information collected using a random stratified sampling design.

A second approach was also used to characterize habitat changes because several map categories had smaller sample sizes in the AA data. To address this sample size issue, we randomly distributed an additional 50 points in each of the seven habitat classes (n = 350) modeled in the 2016 composite habitat map. Although no underwater video or photos were available at these locations, the 2016 composite map had a high overall accuracy (Section 3.4 Composite Habitat Map), and the additional sampling points made this comparison more quantitatively robust across all habitat classes. As was done in the first comparison, underlying habitats from the 2001 IKONOS image were visually interpreted, classified, and scored in a match/mis-match matrix.

While conducting these point-based comparisons, it was observed that a few small areas in the Lagoon experienced major habitat changes clearly visible between the 2001 and 2016 satellite images. These small areas of noteworthy change were not well represented in the Lagoon-wide assessments because points were randomly placed. To capture these changes, we digitized the footprint of the changed areas, plotted them on the 2001 and 2016 images, and qualitatively described the habitat changes based on visual assessment.



Bleached and recently dead Acropora *corals. Credit: M. Kendall and B. Costa.*

3.0 RESULTS AND DISCUSSION

We mapped 27.4 km^2 of seafloor in Saipan Lagoon using BRTs to predict the presence of individual substrate and cover types, and BCTs to generate a composite habitat map. In this section, we present the results from these models, highlight some of the main features of the habitat predictions, convey performance and accuracy of the maps, describe habitat changes in the Lagoon since 2001, and discuss potential applications and ways to customize the products to meet particular research and management needs.

3.1 MODEL PERFORMANCE

Twelve BRT models and resulting spatial predictions describe the probability-of-occurrence for five substrate and seven cover types. Model performance was generally considered ‘good to excellent’ based on four evaluation metrics. Models and predictions with the highest kCV PDE, test PDEs, and AUCs also had the lowest amount of bias and error. For all the models, kCV PDE ranged from 18.1% to 65.9% ($\bar{x}$ = 35.1% ±3.7 SE), and test PDE ranged from -0.63% to 56.6% ($\bar{x}$ = 33.0% ±4.8 SE). The ‘Seagrass (*E. acoroides*)’ model had the highest kCV and test PDEs (65.9% and 56.6%). The ‘Live Coral (Staghorn *Acropora*)’ model had the lowest kCV PDE (18.1%), and the ‘Live Coral (other spp.)’ model had the lowest test PDE (-0.63%). AUC values ranged from 0.73 (good) to 0.94 (excellent; $\bar{x}$ = 0.86 ±0.02 SE) for all the models. The ‘Seagrass (*E. acoroides*)’ model had the highest AUC (0.94), and the ‘Live Coral (other spp.)’ had the lowest (0.73). Bias was small for all models, ranging between -0.06 to +0.11 ($\bar{x}$ = 0.02 ±0.01 SE). Bias indicates whether the model under predicted (-) or over predicted (+) the probability-of-occurrence. ‘Sand’ presence was consistently under-predicted by 6% (e.g., if probability-of-occurrence was 71% in the model, its true value based on AA data was closer to 77%), and ‘Live Coral Reef (all spp.)’ presence was consistently over-predicted (11%). The presence of ‘Seagrass (*E. acoroides*)’, ‘Mixed Algae’, and ‘Live Coral (Staghorn *Acropora*)’ showed no systematic bias. Lastly, RMSE values ranged from 0.1 to 0.42 ($\bar{x}$ = 0.32 ±0.03 SE). The ‘Live Coral Reef (all spp.)’ model had the largest amount of error (0.42), while the ‘Seagrass (*E. acoroides*)’ had the least (0.1).

3.2 PREDICTOR IMPORTANCE

The relative importance of each environmental predictor differed among the substrate and cover models (Figure 3.1). The topographic predictors (i.e., rugosity, slope rate of change) and geographic predictors (i.e., latitude) were the most important drivers behind all of the ‘Live Coral’ models. Distance to reef crest was important for most of the substrate models, including ‘Upright Dead Coral Reef’, ‘Pavement’, ‘Coral Rubble’, and ‘Sand’. Band 12 (Red-Red Edge) also contributed significantly to the ‘Pavement’ model. The remaining cover type models were driven mainly by the spectral predictors, including Band 10 (Green−Red Edge), Band 9 (Green-Red), Band 5 (Coastal Blue−Red Edge) and Band 3 (Blue−Red Edge).

Proxies, such as latitude and distance to reef crest, were important for explaining the distribution of several substrate and cover types. Proxies are predictors that can be easily measured, and may be correlated with other environmental drivers that are challenging to measure. They are often indirectly linked to the underlying ecological mechanism(s) driving habitat distributions. The importance of these proxies indicates that not all of the variability in the habitat data was explained by the environmental predictors used here. Additional information about the environment, such as water temperature or water quality and/or about human uses (e.g., density of use), may be helpful to explain this additional spatial variation. While this information may help identify further research questions, any ecological inference from these results should be used with caution. The models were designed to create the most accurate predictions possible, not to determine which predictors were most ecologically relevant.

Results and Discussion

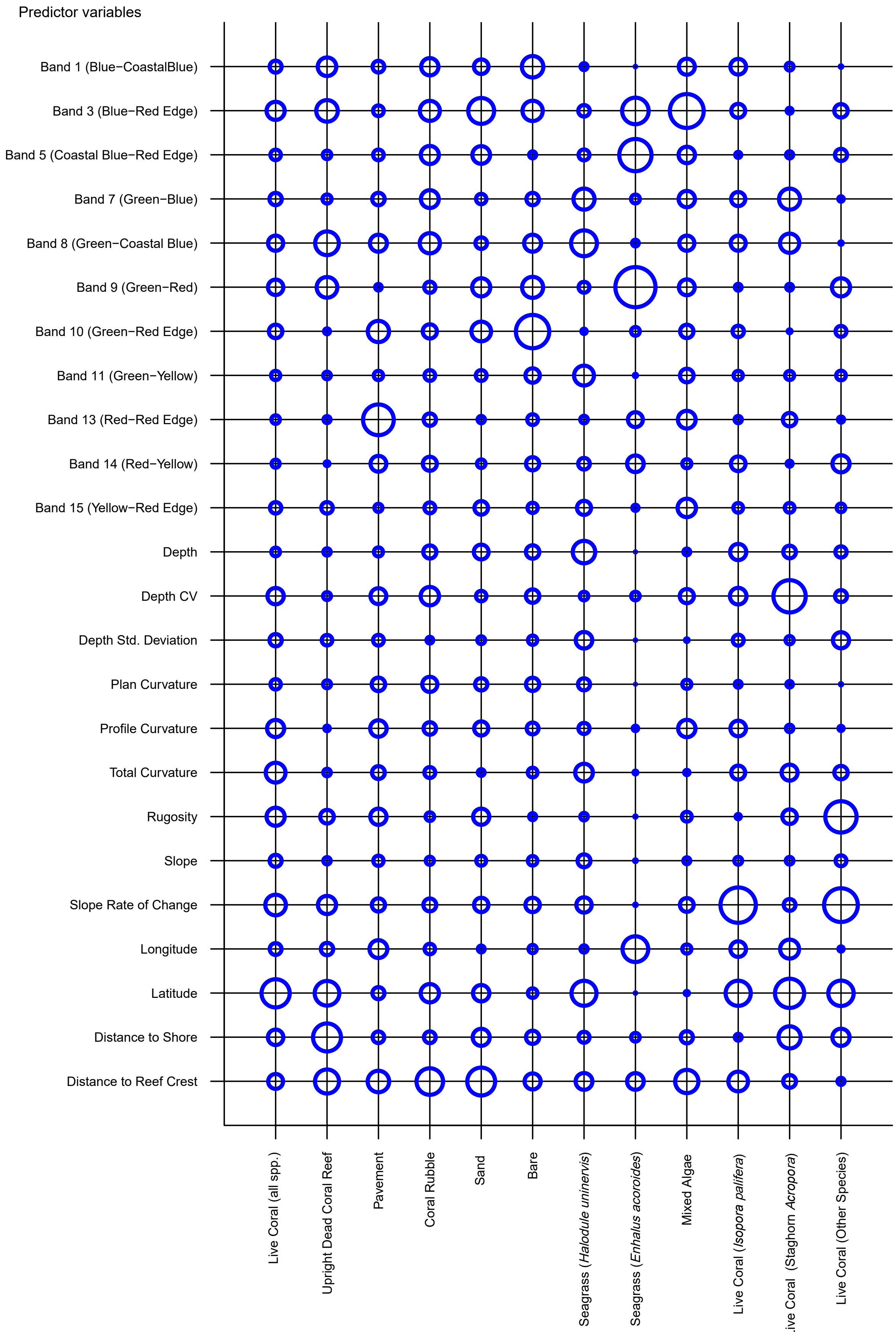


Figure 3.1. Relative importance of the environmental predictors used to develop the 12 habitat models and predictions. Circle size is proportional to a predictor's relative importance averaged across 100 model iterations. The larger the circle, the more important the predictor.

3.3 GEOGRAPHIC PATTERNS OF SUBSTRATE AND COVER TYPES

Substrate: Live Coral Reef (all spp.)

The 'Live Coral Reef (all spp.)' class (Figure 3.2a) was present infrequently in the Lagoon, at only 20% (59/292) of the GV sites (Figure 3.2b). Most observations were north of the channel and inside Mañagaha Marine Conservation Area. The remaining points were clustered mainly along the reef crest north of Tanapag Beach to Wing Beach. The 'Live Coral Reef (all spp.)' model (Figure 3.2c) showed a similar spatial pattern, with the highest likelihood of presence along the reef crest northeast of the harbor, particularly offshore of Pau Pau and Wing Beaches (Figure 3.2d). Probability-of-occurrence values were also high southeast of Mañagaha Island. CV values were lowest (<0.25) in these same locations (Figure 3.2e), indicating higher precision and lower uncertainty for places where live coral is more likely to be present. These spatial patterns broadly match the distributions of live corals depicted in the 2001 map by Houk and van Woesik (2008), including the '*Acropora* and Coral Zone, *Isopora* and *Porites* (sparse and thick), Live mixed coral back reef flat, and *Porites* and mixed coral zone' habitat classes. It also broadly matched the 'Coral 10%-<50%' class in the 2001 NOAA map (NOAA NCCOS, 2005). Two notable exceptions are areas along the channel and harbor, where more live coral was observed and predicted in 2016.

Substrate: Upright Dead Coral Reef

The 'Upright Dead Coral Reef' class (Figure 3.3a) was fairly common and was present at 30% (89/292) of the GV sites (Figure 3.3b). However, this class was somewhat spatially clustered, often being more abundant and more frequently present between Tanapag and Wing Beaches, and inside the Mañagaha Marine Conservation Area. 'Upright Dead Coral Reef' was also present in higher abundances along the southern edge of the channel, and south of the Lighthouse Reef Trochus Sanctuary. The model (Figure 3.3c) shows similar spatial patterns, with the highest likelihood of presence along the reef crest from Wing Beach to the channel, and from the Lighthouse Reef Trochus Sanctuary to the Sugar Dock (Figure 3.3d). Probability-of-occurrence values were also high about one kilometer north of Agigan Point. CV values were lowest (<0.25) in these same locations (Figure 3.3e), indicating higher precision and lower uncertainty for places where dead coral reef is more likely to be present. These spatial patterns broadly match the distributions of 'Aggregate Reef' in the 2001 NOAA map (NOAA NCCOS, 2005). However, it is difficult to determine how they compare to the 2001 Houk map (Houk and van Woesik, 2008) because of the differences between classification schemes.

Substrate: Pavement

The 'Pavement' class was fairly common in the Lagoon and was present at 30% (87/292) of the GV sites (Figure 3.4b). 'Pavement' was also evenly distributed throughout the Lagoon both north and south of the channel. However, it was slightly more abundant south of the channel, particularly from southern Garapan to Agingan Point. The 'Pavement' model (Figure 3.4c) showed similar spatial patterns, with the highest likelihood of presence on many reef crests and in the Lagoon between Susupe Point south to Agingan Point (Figure 3.4d). Probability-of-occurrence values were also high inside the Mañagaha Marine Conservation Area and from Tanapag Beach north to Wing Beach. CV values were lowest (<0.25) in these same locations (Figure 3.4e), indicating higher precision and lower uncertainty for places where pavement is more likely to be present. These spatial patterns differ from the 'Pavement' distributions in the 2001 NOAA map (NOAA NCCOS, 2005), most likely because of the coarser scale of the map. It is difficult to determine how they compare to the 2001 Houk map (Houk and van Woesik, 2008) because of the differences between classification schemes.

Results and Discussion

Substrate: Live Coral, All Species

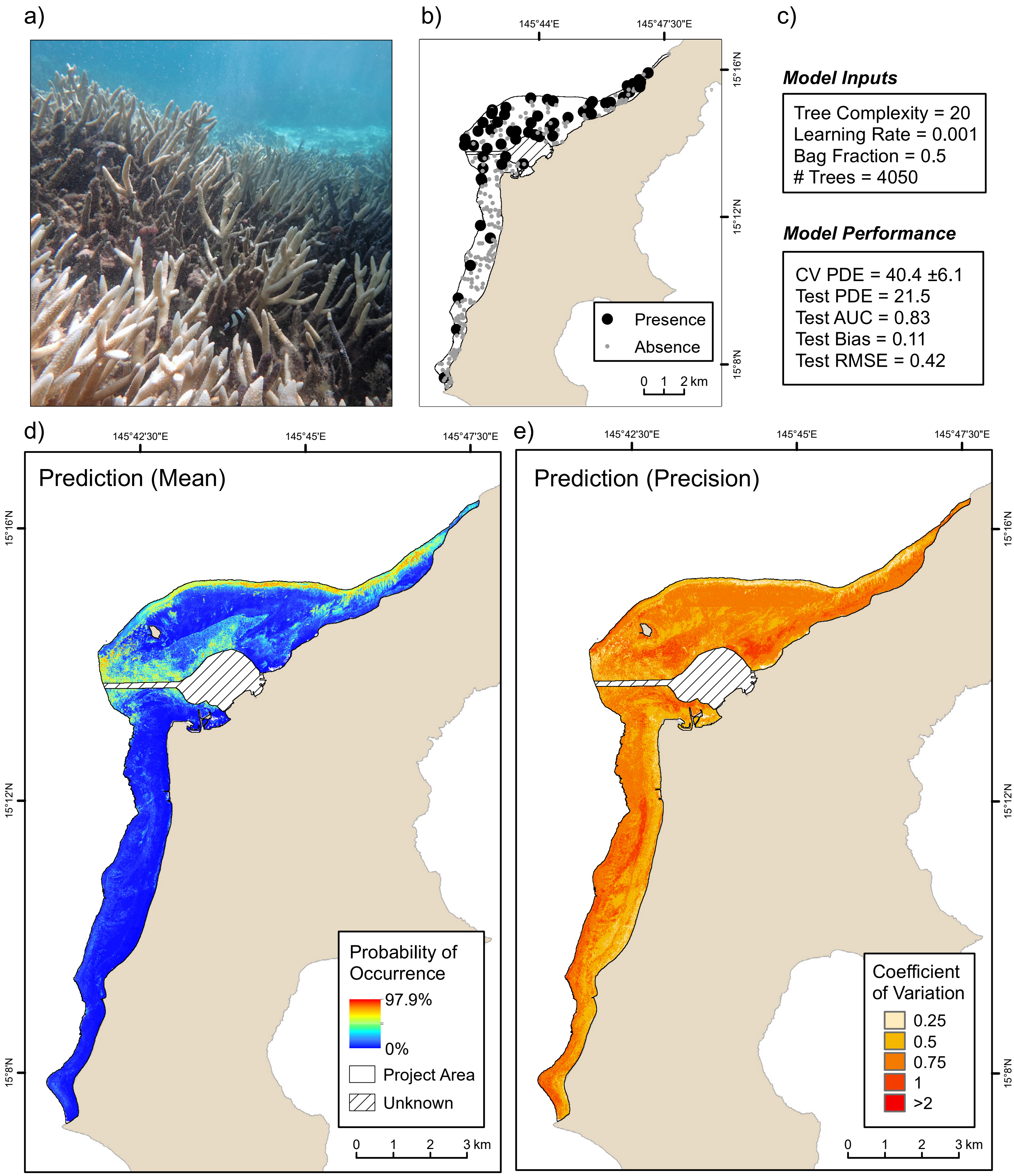


*Figure 3.2. Predicted presence of 'Live Coral Reef (all spp.)'. Figure panels depict: a) a photo of live coral (*Acropora spp.*); b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.*

Substrate: Upright Dead Coral Reef

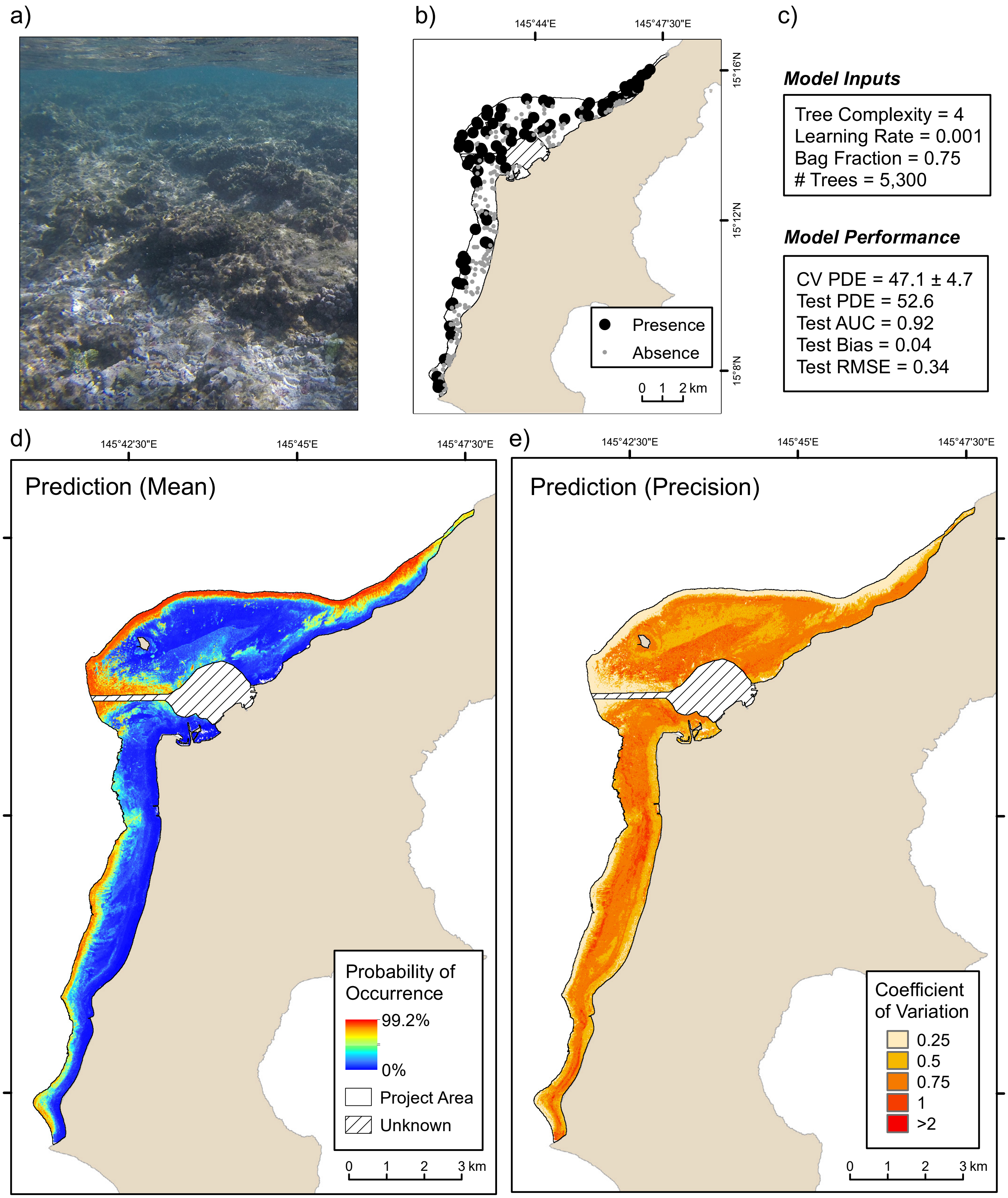


Figure 3.3. Predicted presence of 'Upright Dead Coral Reef'. Figure panels depict: a) a photo of upright dead coral reef; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; e) the predicted average probability-of-occurrence and f) coefficient of variation for the model.

Substrate: Pavement

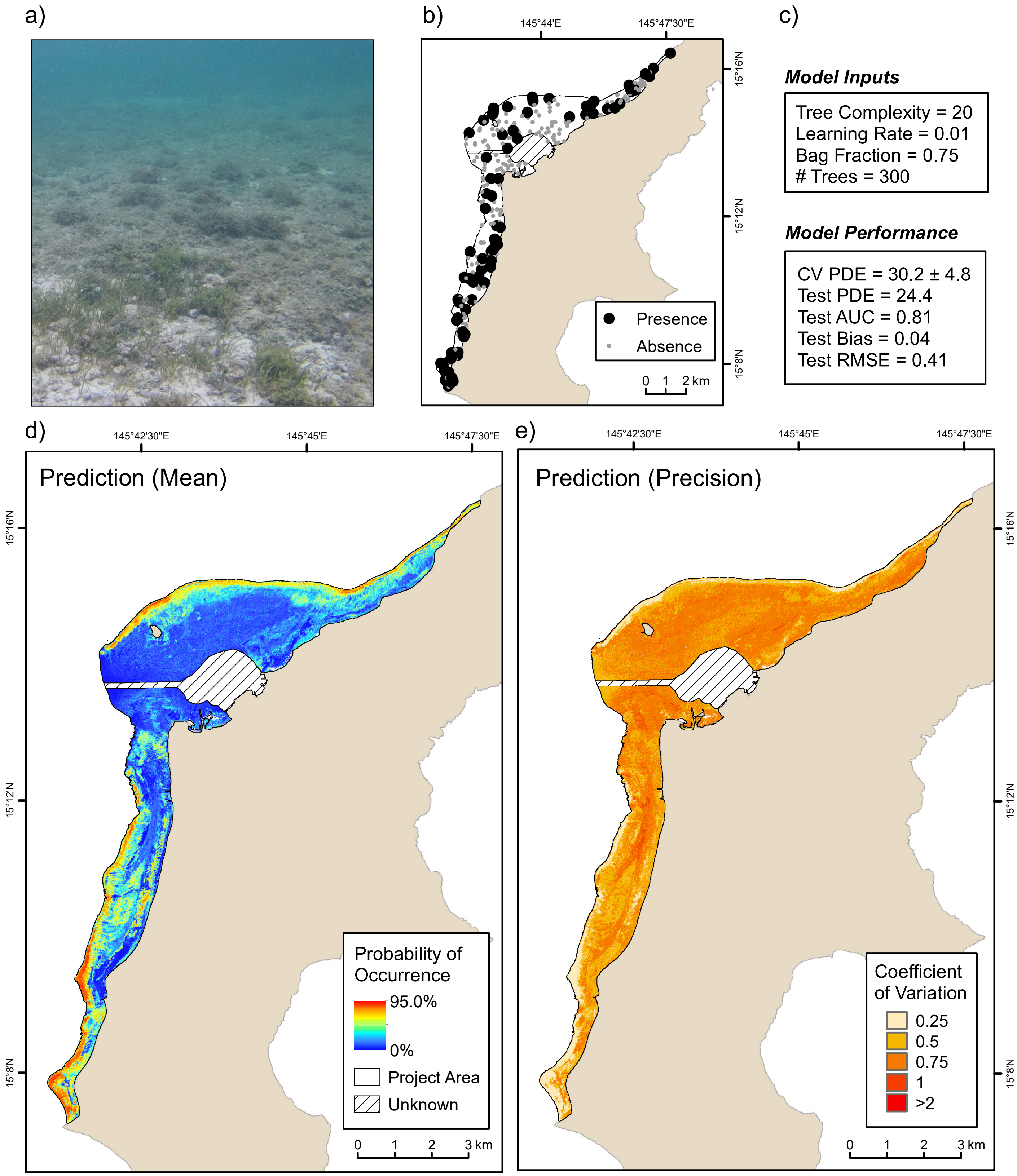


Figure 3.4. Predicted presence of 'Pavement'. Figure panels depict: a) a photo of pavement; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.

Substrate: Coral Rubble

The 'Coral Rubble' class (Figure 3.5a) was common in the Lagoon. Specifically, it was present at 39% (113/292) of the GV sites that we visited (Figure 3.5b). This class was also rather evenly distributed behind the reef crest, except for areas west of Garapan (where there was no defined reef crest), and close to shore between Susupe Point and Oleai. This habitat was also slightly more abundant directly north of Garapan (just south of the harbor). The 'Coral Rubble' model (Figure 3.5c) showed similar spatial patterns, with the highest likelihood of presence just inshore of the defined reef crest, and then scattered shoreward in lower values (Figure 3.5d). Like the other models, CV values were lowest (<0.25) in these same locations (Figure 3.5e), indicating higher precision and lower uncertainty for places where coral rubble is more likely to be present. This habitat is predicted to be present in more locations (particularly to the west and north of Mañagaha Island) than depicted by the 'Coral Rubble' class in the 2001 NOAA map (NOAA NCCOS, 2005). These two maps differ most likely because of the coarser scale of the NOAA map. It is difficult to determine how they compare to 2001 Houk map (Houk and van Woesik, 2008) because of the differences between classification schemes.

Substrate: Sand

The 'Sand' class (Figure 3.6a) was very common shoreward of the reef crest. It was present at 67% (196/292) of the GV sites (Figure 3.6b), and was evenly distributed throughout the Lagoon. However, it had notably lower abundances along the reef crest from Mañagaha Island to Wing Beach, southeast of Oleai and up to 1.5 km north of Agingan Point. The 'Sand' model (Figure 3.6c) showed similar spatial patterns, with the highest likelihood of presence approximately 125 m inshore of the defined reef crest (Figure 3.6d). Like the other models, CV values were lowest (<0.25) in these same locations (Figure 3.6e), indicating higher precision and lower uncertainty. These spatial patterns broadly match the distributions of 'Sand' and 'Sand with Scattered Coral and Rock' classes in the 2001 NOAA map (NOAA NCCOS, 2005). The notable exception is the area between Tanapag and Pau Pau Beaches, where sand is predicted to be present in 2016. More sand is also predicted in most locations around the Lagoon compared to the 'Sand dominated coastal shallow depths', 'Sand dominated mid Lagoon depths' and various seagrass classes in the 2001 Houk map (Houk and van Woesik, 2008). These two maps differ most likely because of the coarser scale of the earlier map.

Cover: Live Coral (*Isopora palifera*)

The 'Live Coral (*Isopora palifera*)' class (Figure 3.7a) was rare in the Lagoon, occurring at only 5% (16/292) of the GV sites (Figure 3.7b). This species occurred more frequently north of Tanapag. Approximately half (17/39) of the occurrences were inside the Mañagaha Marine Conservation Area. The remaining occurrences were clustered offshore of Pau Pau Beach and northwest of Tanapag Beach. The area northwest of Tanapag Beach also had some of the highest abundances. The model (Figure 3.7c) showed similar spatial patterns, with the highest likelihood of *I. palifera* presence about 200 m shoreward of the reef crest north of Tanapag (Figure 3.7d). Probability-of-occurrence values were also high offshore of Pau Pau Beach. Unlike the previous models, the CV values were lowest (<0.25) where the model predicted low probabilities-of-occurrence and highest (>0.75) where it was more likely to be present (Figure 3.7e). This inverted pattern indicates that there is lower uncertainty for locations where *I. palifera* is likely to be absent, and higher uncertainty for locations where it is likely to be present. This inverted pattern is due to the rarity of this coral. Spatial patterns seen in the probability-of-occurrence prediction broadly match the distributions of '*Isopora* and *Porites* (Sparse) and (Thick)' classes in the 2001 Houk map (Houk and van Woesik, 2008). Two notable exceptions are areas along the reef crest west and south of Mañagaha Island, where more *I. palifera* was recorded and is predicted to be present in 2016. It is difficult to determine how these patterns compare to the 2001 NOAA map (NOAA NCCOS, 2005) because *I. palifera* was not explicitly mapped.

Results and Discussion

Substrate: Coral Rubble

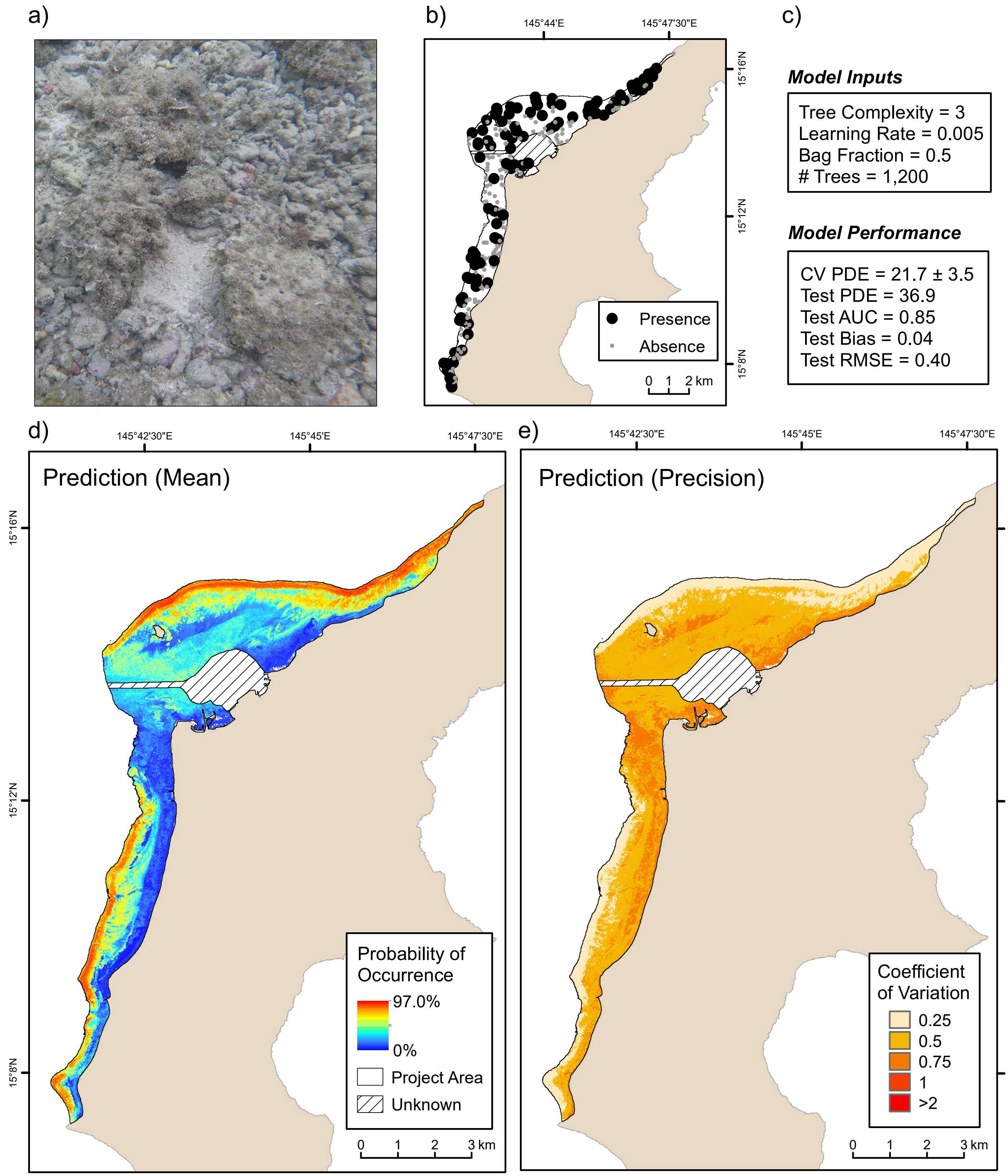


Figure 3.5. Predicted presence of 'Coral Rubble'. Figure panels depict: a) a photo of coral rubble; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the 'model.

Substrate: Sand

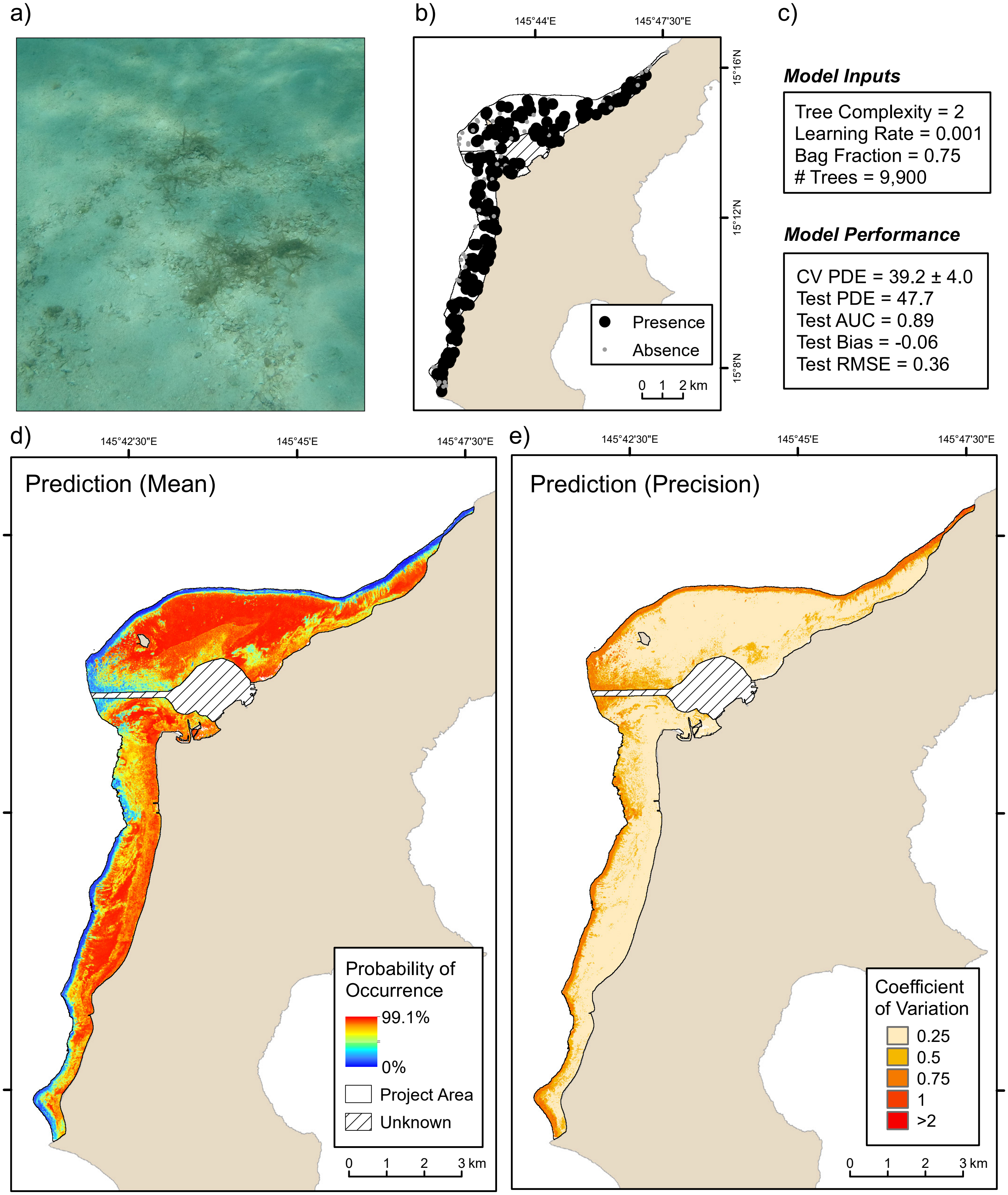


Figure 3.6. Predicted presence of 'Sand'. Figure panels depict: a) a photo of sand; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.

Cover: Live Coral, *Isopora palifera*

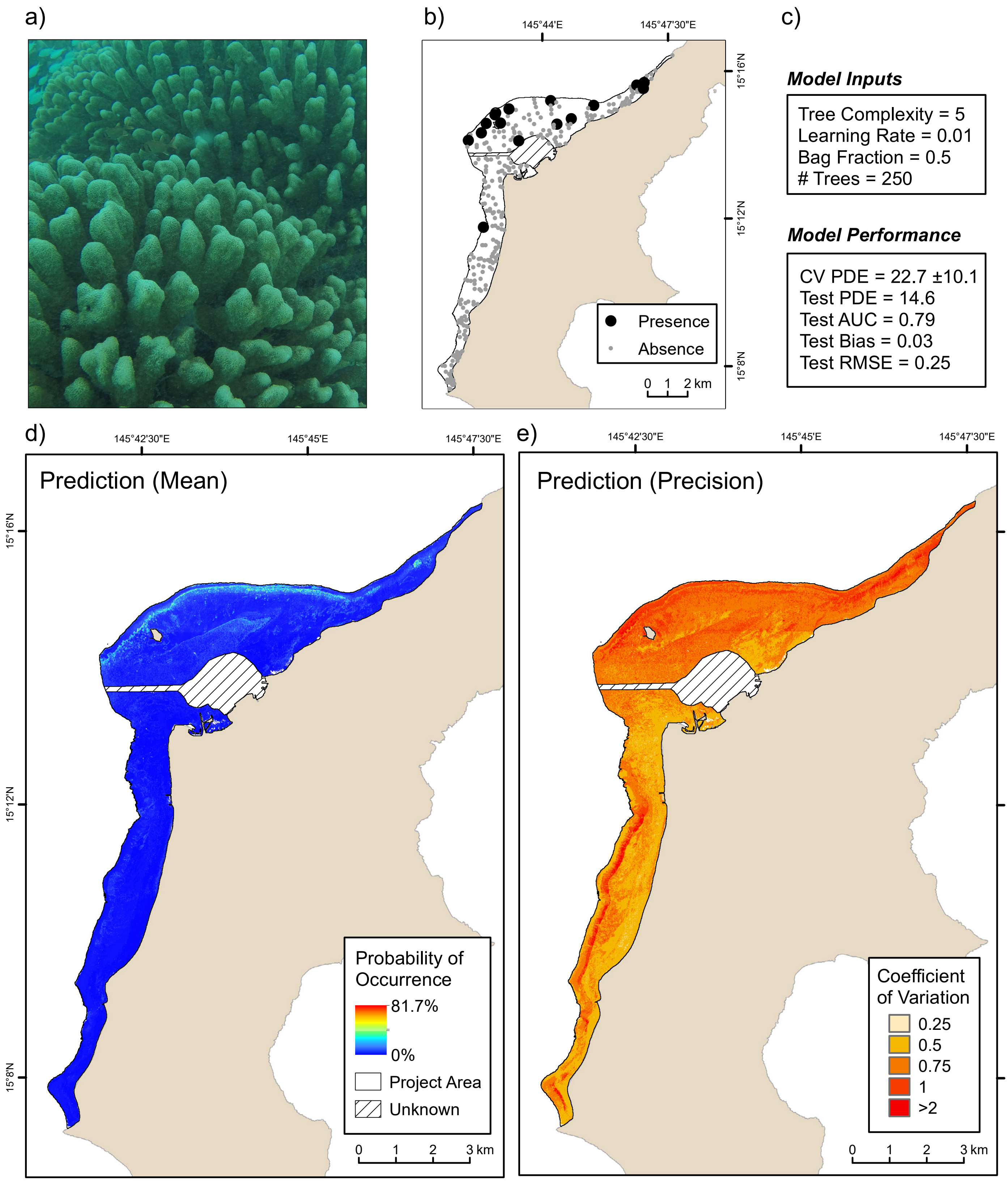


*Figure 3.7. Predicted presence of 'Live Coral (*Isopora palifera*)'. Figure panels depict: a) a photo of* I. palifera*; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.*

Cover: Live Coral (Staghorn *Acropora*)

'Live Coral (Staghorn *Acropora*)' (Figure 3.8a) was very rare in the Lagoon, occurring at only 4% (13/292) of the GV sites (Figure 3.8b). These three Staghorn *Acropora* species occurred more frequently north of Tanapag, with over half (8/13) of the occurrences offshore of Pau Pau Beach. The remaining sitings were sparsely distributed south of the channel, mainly offshore of Oleai. The model (Figure 3.8c) showed similar spatial patterns, with the highest likelihood of these species being present about 600 m offshore of Pau Pau Beach (Figure 3.8d). The areas approximately 1.1 km north of Agingan Point, and 600 m offshore of Susupe Point, also had slightly higher probabilities-of-occurrence. Like with the model for *I. palifera*, the CV values were lowest (<0.5) for the Live Coral (Staghorn *Acropora*) model where it predicted low probabilities-of-occurrence, and highest (>1) where it was more likely to be present (Figure 3.8e). This inverted pattern is due to the rarity of these three *Acropora* species, and it indicates that there is lower model uncertainty for locations where these species are likely to be absent, and higher uncertainty for locations where they are likely to be present. It is difficult to determine how these patterns compare to the 2001 NOAA map (NOAA NCCOS, 2005) because these species were not explicitly mapped in that study. However, spatial patterns seen in the probability-of-occurrence prediction broadly match the distributions of '*Acropora* and Coral Zone' class in the 2001 Houk map (Houk and van Woesik, 2008). Two notable exceptions are areas inside and north of Lighthouse Reef Trochus Sanctuary, where more was mapped in the 2001 imagery.

Cover: Live Coral (other spp.)

The 'Live Coral (other spp.)' class (Figure 3.9a) includes all live coral species except for *I. palifera*, Staghorn *Acropora*, and massive *Porites* species. Like other live coral groups, these species were present infrequently, occurring at only 11% (31/292) of the GV sites (Figure 3.9b). They were present more frequently north of the channel, with about a third (23/84) of their presences and half (3/6) of their highest abundances occurring inside Mañagaha Marine Conservation Area. They were also concentrated between Pau Pau and Wing Beaches and offshore of Tanapag Beach. The remaining presences were sparsely distributed south of the channel, mainly offshore of Oleai and north of Agingan Point. This model (Figure 3.9c) showed similar spatial patterns, with the highest likelihood of presence along the seaward edges of the channel (Figure 3.9d). A few other areas also had higher probabilities-of-occurrence, including the reef crest from Tanapag to Wing Beaches, areas north and west of Garapan, and areas southwest of the Sugar Dock. Unlike some of the other live coral models, the CV values were lowest (<0.5) where the model predicted the highest and lowest probabilities-of-occurrence. CV values were highest (>1) for moderate probability-of-occurrence values (Figure 3.9e). This bimodal pattern indicates that there is lower uncertainty for locations where these live coral species are very likely to be absent or present. It is difficult to determine how these patterns compare to the maps based on 2001 imagery because species other than *I. palifera*, *Acropora formosa*, *Acropora aspera*, and massive *Porites* species were not explicitly mapped by Houk and van Woesik (2008) and NOAA NCCOS (2005).

Cover: Seagrass (*Enhalus acoroides*)

The 'Seagrass (*Enhalus acoroides*)' class (Figure 3.10a) was only present at 5% (15/292) of the GV sites (Figure 3.10b). However, it formed dense patches when it was present, particularly very near shore (<250 m) between Tanapag Beach and Oleai, and near Tanapag, Smiling Cove and shoreward of the Lighthouse Reef Trochus Sanctuary. The model (Figure 3.10c) showed similar spatial patterns, with the highest likelihood of presence between Tanapag Beach and Oleai (Figure 3.10d). A few other areas also had higher probabilities-of-occurrence, including isolated patches just south of Oleai and around Susupe and Agingan Points. CV values were lowest (<0.25) in these same locations (Figure 3.10e), indicating higher precision and lower uncertainty for places where *E. acoroides* is likely to be present. The spatial patterns seen in the probability-of-occurrence prediction closely match the distributions of the '*H. macroloba*, *Halodule*, *Enhalus* and Thick *Enhalus* Zone' classes in the 2001 Houk map (Houk and van Woesik, 2008). The notable exceptions are the isolated patches just south of Oleai and around Susupe and Agingan Points, where *E. acoroides* is predicted to be present in 2016. It is difficult to determine how these patterns compare to the 2001 NOAA map (NOAA NCCOS, 2005) because *E. acoroides* was not explicitly mapped.

Results and Discussion

Cover: Live Coral, Staghorn (*Acropora aspera, A. formosa and A. pulchra*)

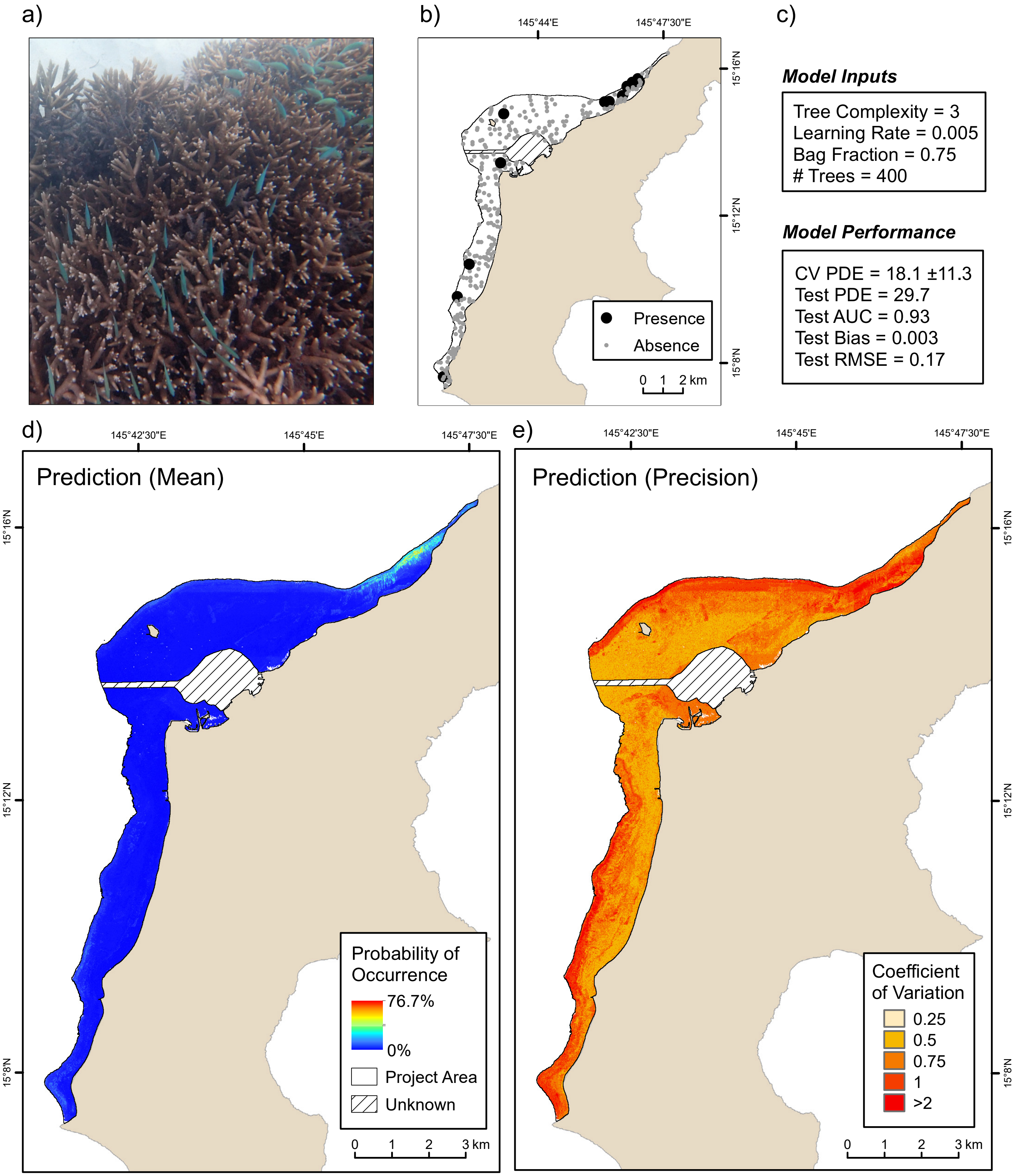


Figure 3.8. Predicted presence of 'Live Coral (Staghorn Acropora*)'. Figure panels depict: a) a photo of* Acropora *spp.; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.*

Cover: Live Coral, Other Species

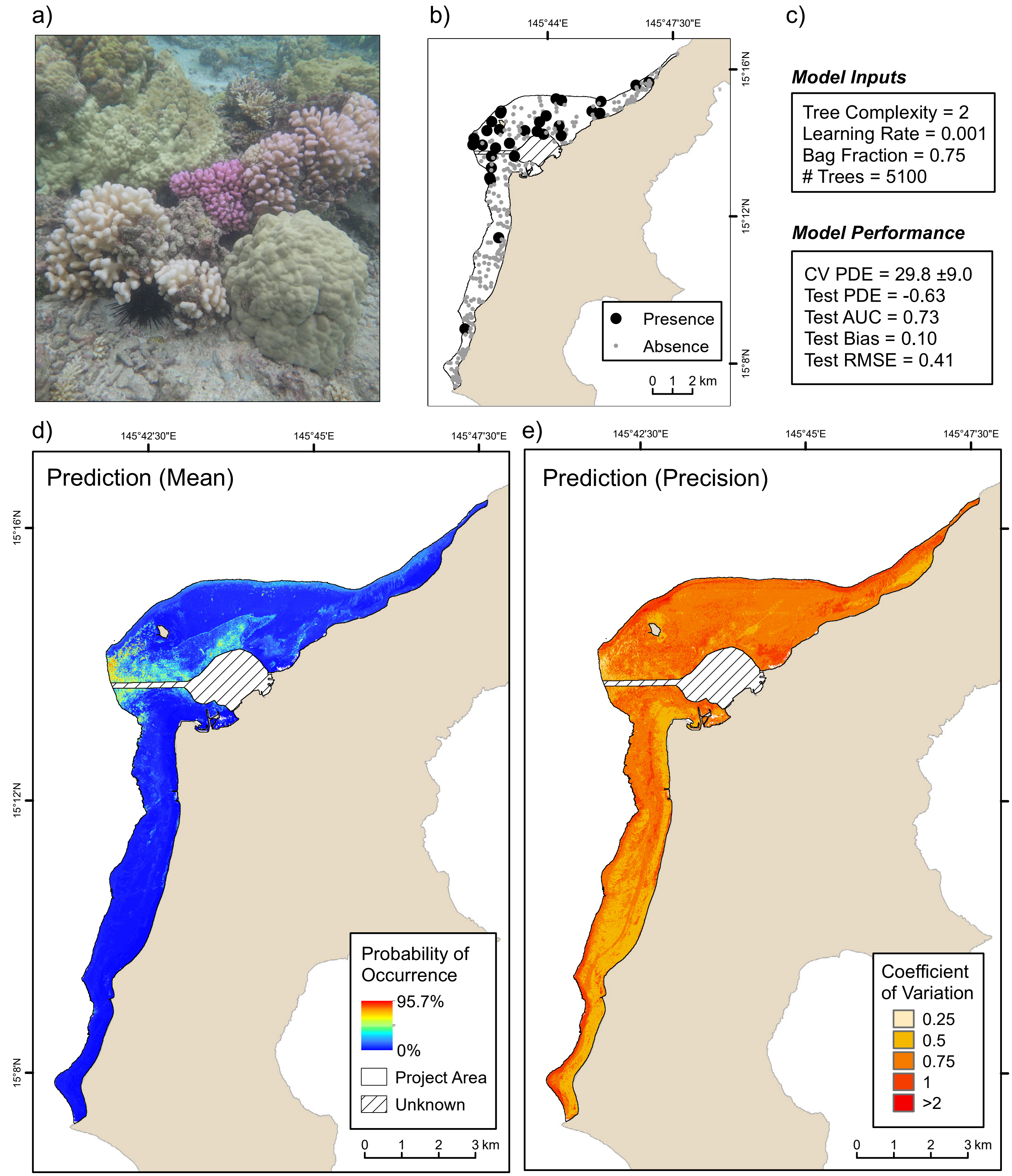


Figure 3.9. Predicted presence of 'Live Coral (other spp.)'. Figure panels depict: a) a photo of multiple hard coral species; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.

Results and Discussion

Cover: Seagrass, *Enhalus acoroides*

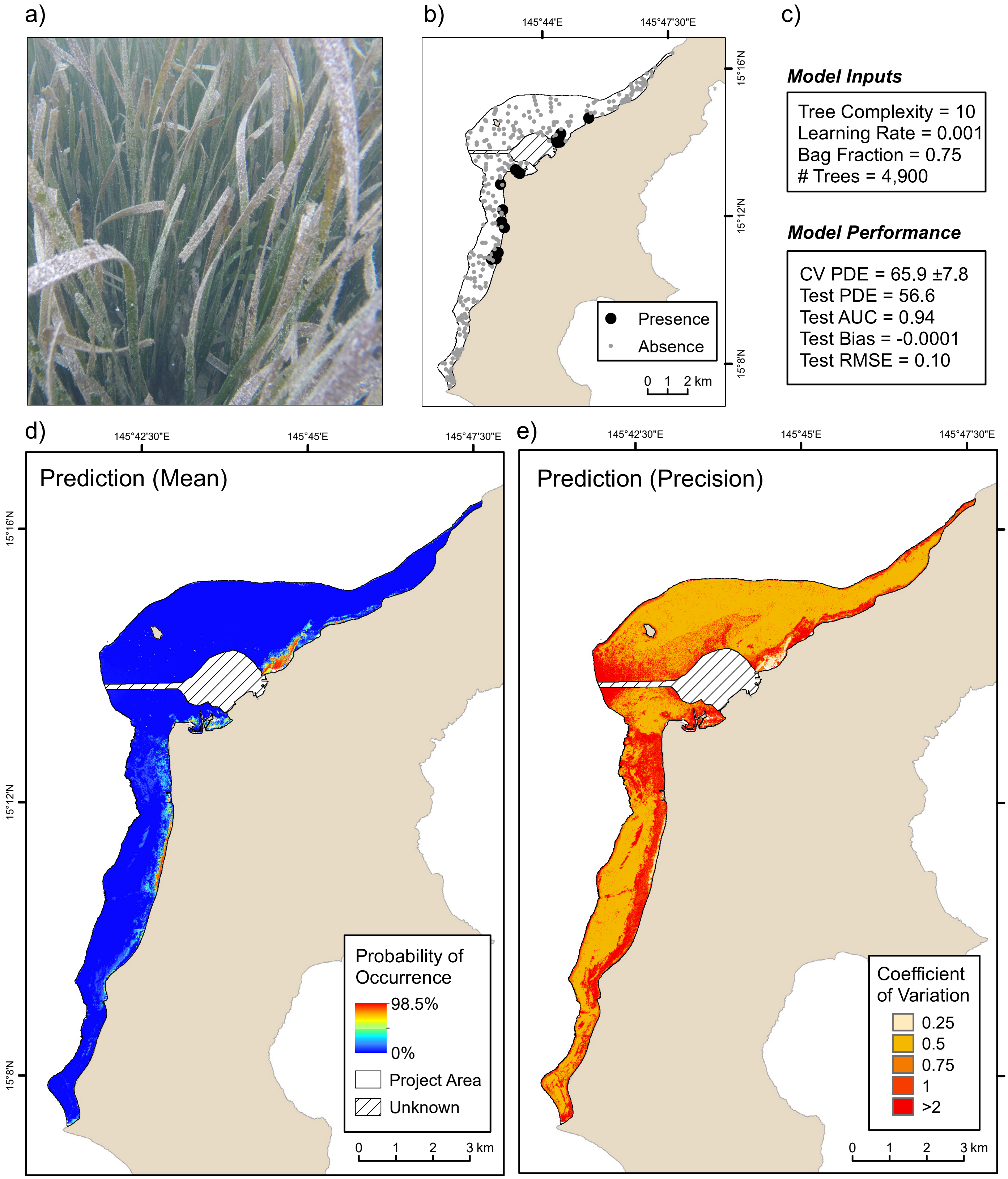


*Figure 3.10. Predicted presence of 'Seagrass (*Enhalus acoroides*)'. Figure panels depict: a) a photo of* E. acoroides*; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.*

Cover: Seagrass (*Halodule uninervis*)

The 'Seagrass (*Halodule uninervis*)' class (Figure 3.11a) was present at 28% (83/292) of the GV sites (Figure 3.11b). It was spatially clustered, occurring more frequently and in higher abundances from Garapan south to Agingan Point. It was more sparsely present from Tanapag north to Pau Pau Beach. In both locations, *H. uninervis* was often < 500 m from the shoreline. The model (Figure 3.11c) showed similar spatial patterns, with the highest likelihood of being present < 500 m from shore between Garapan and Agingan Point (Figure 3.11d). A few other areas also had higher probabilities-of-occurrence, including isolated patches from Tanapag north to Pau Pau Beach. CV values were lowest (<0.25) south of Garapan (Figure 3.11e), indicating lower uncertainty for places in the southern Lagoon where this seagrass species is likely to be present. Conversely, CV values were higher (>0.5) for locations with low probabilities-of-occurrence south of Garapan, and with higher probabilities-of-occurrence north of Garapan. This bimodal pattern indicates that there is higher uncertainty for locations in the southern Lagoon where *H. uninervis* are very likely to be absent, and for locations in the northern Lagoon on the whole. The spatial patterns seen in the probability-of-occurrence prediction closely match the distributions of the '*H. macroloba*, *Halodule*, *Enhalus*, *Halodule* bunkers, *Halodule* mix (zone), and Rocky *Halodule* (zone)' classes in the 2001 Houk map (Houk and van Woesik, 2008). The notable exceptions are the areas < 500 m offshore of Oleai and areas west of Garapan, where *E. acoroides* is predicted to be present in 2016. It is difficult to determine how these patterns compare to the 2001 NOAA map (NOAA NCCOS, 2005) map because *H. uninervis* was not explicitly mapped.

Cover: Mixed Algae

The 'Mixed Algae' class (Figure 3.12a) was common and evenly distributed throughout the Lagoon. 'Mixed Algae' habitats were present at 82% (242/292) of the GV sites (Figure 3.12b). The only sizable area (approximately 1 km^2) without this class was located approximately two kilometers west of Tanapag Beach. The model (Figure 3.12c) showed similar spatial patterns, with the lowest likelihood of this habitat being present close to shore near Tanapag, about one kilometer west of Smiling Cove and along a swath of seafloor west of Mañagaha Island (Figure 3.12d). Probability-of-occurrence for 'Mixed Algae' was high in most other locations. CV values were lowest (<0.25) for the model in places with high probabilities-of-occurrence (Figure 3.12e), indicating lower uncertainty for places where it is very likely to be present. The spatial patterns seen in the probability-of-occurrence prediction are quite different than the distributions of the 'Heavy Macroalgae Zone, Macroalgae back reef flat and Pleistocene rock macro *Halodule* patch' classes in the 2001 Houk map (Houk and van Woesik, 2008). They also do not match the 'Coralline Algae, Macroalgae and Turf' classes in the 2001 NOAA map (NOAA NCCOS, 2005). In both cases, the 'Mixed Algae' model predicts these habitats are likely to be present in many more locations than in the earlier maps based on 2001 imagery. The distribution of 'Mixed Algae' most likely differs widely in these maps because of differences in the scales and classification schemes of the various maps.

Cover: Bare

Locations without biological cover (i.e., 'Bare'; Figure 3.13a) were common and evenly distributed throughout the Lagoon. The 'Bare' cover class was present at 54% (159/292) of the GV sites (Figure 3.13b). The model (Figure 3.13c) showed similar spatial patterns, with the lowest likelihood of this habitat being present along the reef tract, close to shore near Tanapag, and in isolated patches between Garapan and Agingan Point (Figure 3.13d). Probability-of-occurrence for this class was high in most other locations inside the Lagoon. CV values were lowest (<0.25) in places with high probabilities-of-occurrence (Figure 3.13e), indicating lower uncertainty for places where it is very likely to be present. It is difficult to determine how the spatial patterns seen in the probability-of-occurrence prediction compare to the 2001 Houk map (Houk and van Woesik, 2008), because 'Bare' cover was not explicitly mapped. However, these spatial distributions are quite different than the distributions of the 'Uncolonized 90%-100%' class in the 2001 NOAA map (NOAA NCCOS, 2005). The 'Bare' model predicts the presence of bare substrate in many more locations than in 2005, including between Garapan and Susupe Point, and from Mañagaha Island around the reef tract to Pau Pau Beach. The distribution of bare bottom differs widely between the 2001 NOAA and 2016 maps because of the much coarser scale used by NOAA to delineate habitats (i.e., 4,047 m^2 in 2001 versus 4 m^2 in 2016).

Cover: Seagrass, *Halodule uninervis*

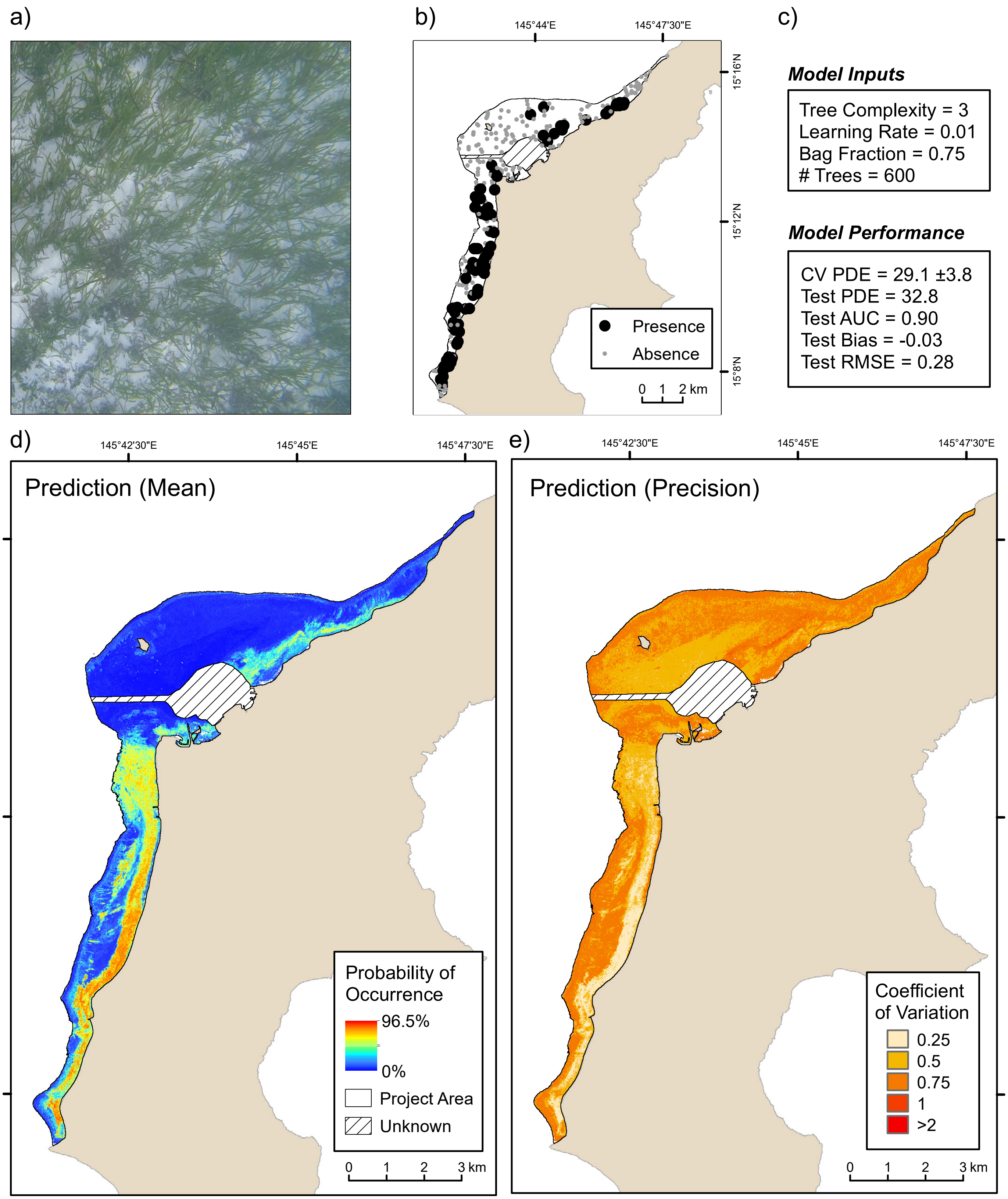


*Figure 3.11. Predicted presence of 'Seagrass (*Halodule uninervis*)'. Figure panels depict: a) a photo of* H. uninervis*; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.*

Cover: Mixed Algae

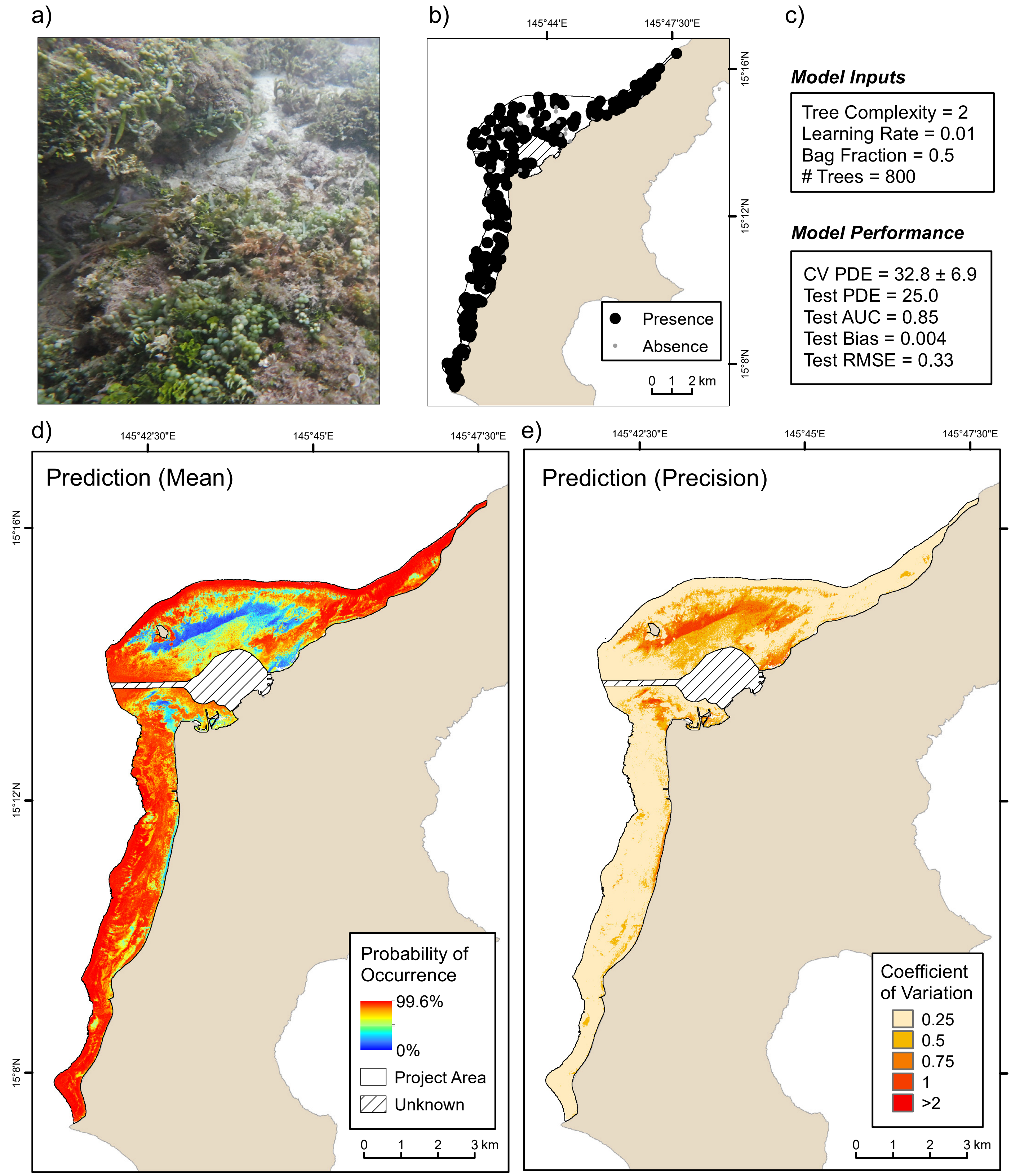


Figure 3.12. Predicted presence of 'Mixed Algae'. Figure panels depict: a) a photo of mixed algal habitat; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.

Cover: Bare

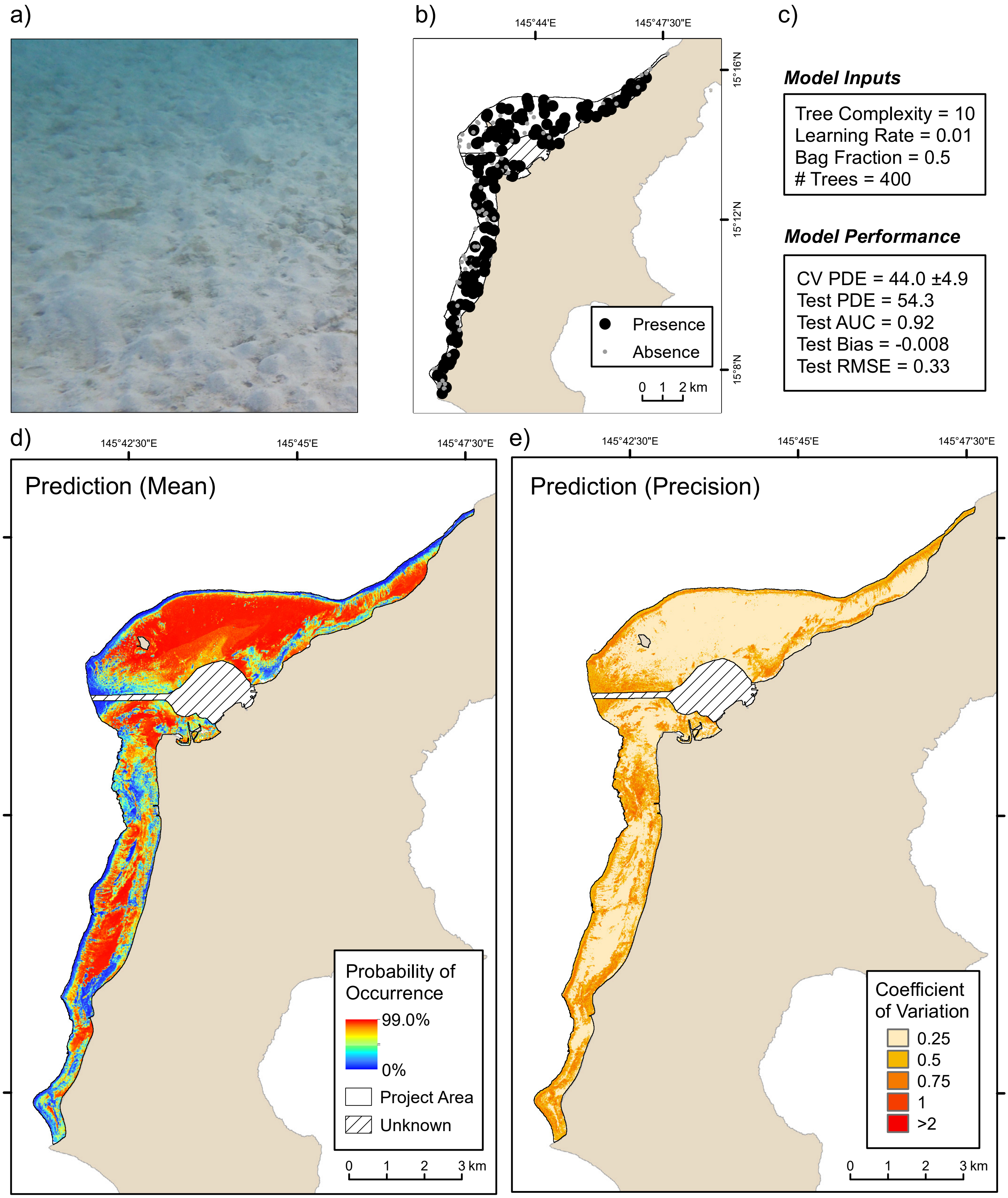


Figure 3.13. Predicted presence of 'Bare' cover. Figure panels depict: a) a photo of mixed bare cover habitat; b) a map denoting the presences and absences of this class in the GV data; c) the input parameters used to create the final model and the performance of the final model; d) the predicted average probability-of-occurrence and e) coefficient of variation for the model.

3.4 COMPOSITE HABITAT MAP

We characterized approximately 27.4 km^2 of seafloor in Saipan Lagoon using BCTs. This composite habitat map displays the predicted spatial distribution of seven common combinations of substrate and cover types (Figure 3.14). 'Bare Sand' was the most abundant habitat type mapped, comprising 32.6% (8.9 km^2) of the area. The largest, continuous patches of 'Bare Sand' were located north of Garapan and east of Mañagaha Island. There were also patches between Susupe Point and Oleai. 'Pavement Colonized with Mixed Algae' was the next most abundant habitat mapped, comprising 25.7% (7.1 km^2) of the area. Large, continuous patches of 'Pavement Colonized with Mixed Algae' were most common between Garapan and Agingan Point. Smaller patches were also located along the reef tract from Mañagaha Island north to Wing Beach, along with patches of 'Live or Upright Dead Coral Colonized with Mixed Algae'. This habitat type was the third most abundant habitat mapped in the Lagoon, comprising 13.9% (3.81 km^2) of the habitat map. It was also present in large, continuous patches along the seaward edges of the channel.

'Mixed Algae and Seagrass on Sand' was the fourth most abundant habitat, comprising 13.8% (3.79 km^2) of the area. Its distribution was more diffuse than the other habitats, particularly between Garapan and Sugar Dock. However, there were some larger, more continuous patches offshore of Tanapag Beach north to Pau Pau Beach. 'Seagrass (*H. uninervis*) on Sand' was the fifth most abundant habitat mapped, comprising 7.1% (2.0 km^2) of the area. Large, continuous beds of *H. uninervis* were most common between Garapan and Sugar Dock. There were some more diffuse beds in nearshore areas around Smiling Cove and between Tanapag and Tanapag Beach. 'Coral Rubble Colonized with Mixed Algae' was the sixth most abundant habitat, comprising 4.1% (1.1 km^2) of the area. The most continuous patches of this habitat were located about 100 m inshore of the reef crest between Garapan and Sugar Dock. Smaller patches were also mapped off of Pau Pau Beach.

'Seagrass (*E. acoroides*) on Sand' was the least abundant habitat mapped in the Lagoon, comprising just 2.7% (0.7 km^2) of the area. There were three main spatial clusters of this seagrass species all of which were very nearshore. The largest and densest cluster was within about 500 m of the Tanapag shoreline. The second, smaller and more diffuse cluster was located near Smiling Cove Marina. The majority of the third cluster was located between Garapan and Oleai. Beds of *E. acoroides* were much less dense between Garapan and the northern boundary of the Lighthouse Reef Trochus Sanctuary. These beds became more dense between the northern boundary of the Lighthouse Reef Trochus Sanctuary and Oleai.

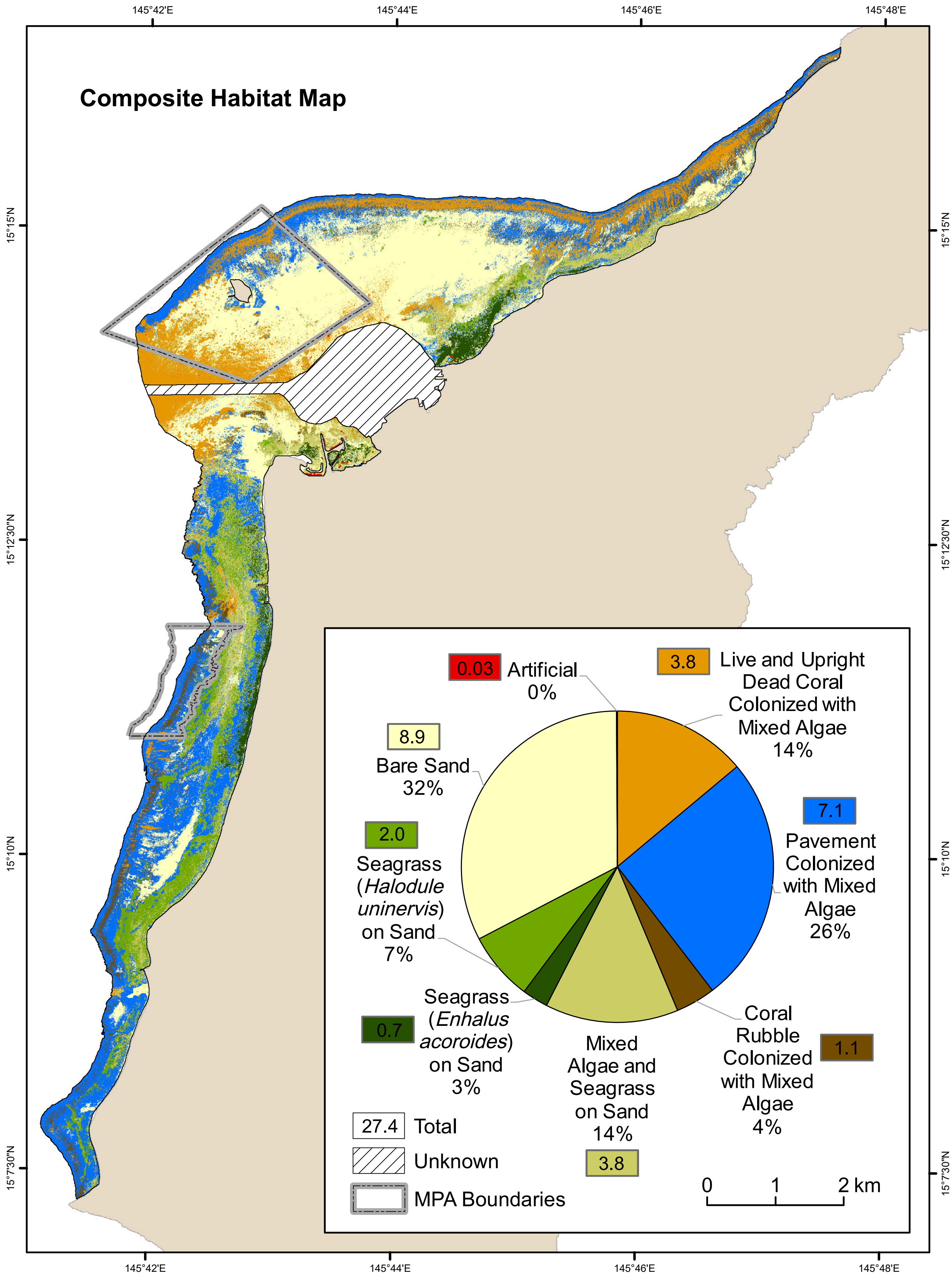


Figure 3.14. This figure depicts the seven benthic habitats (plus 'Artificial') mapped throughout Saipan Lagoon. The numbers inside the legend denote the amount of area (km^2) occupied by each habitat class.

Map Accuracy

The relative prevalence and proportions of the seven habitat types were very similar in both the field data and the final composite habitat map. Agreement between the field data and habitat map suggests the BCTs were able to describe the relationships among the habitats and environmental predictors reasonably well. The overall accuracy and tau value for the composite habitat map (quantified using the AA points) was high, at 85.7% and 0.83 ±0.06, respectively (Table 3.1). The user's accuracies were also high, ranging from 80% to 100% for the individual habitat classes. The overall accuracy was very similar after correcting for proportional biases (85.8% ±4.5% at the 95% confidence level). Habitat misclassifications were evenly distributed in the confusion table, suggesting that the habitat characterization process did not consistently confuse any particular pair of substrate and cover types. These accuracies are similar to the other benthic habitat maps created by NOAA NCCOS in the Pacific Region (NOAA NCCOS, 2005; Battista et al. 2007). As a result, this habitat map can be used with high levels of confidence for a variety of research and management applications.

Table 3.1. The confusion matrix for the composite habitat map. Accuracy Assessment sites are listed as columns and corresponding mapped habitats, as rows. Cell values are the number of matches (along the diagonal) or mismatches (off diagonal) between the two.

AA (i)

Map (i)	Mixed Algae & Seagrass on Sand	Seagrass (*Halodule uninervis*) on Sand	Bare Sand	Seagrass (*Enhalus acoroides*) on Sand	Pavement Colonized with Mixed Algae	Coral Rubble Colonized with Mixed Algae	Live & Upright Dead Coral Colonized with Mixed Algae	n_j	User's Accuracy (%)
Mixed Algae & Seagrass on Sand	**26**				1	2	1	**30**	87%
Seagrass (*Halodule uninervis*) on Sand		**11**			2			**13**	85%
Bare Sand	2	1	**51**		4	2	4	**64**	80%
Seagrass (*Enhalus acoroides*) on Sand				**5**				**5**	100%
Pavement Colonized with Mixed Algae	2	2	1		**75**	2	5	**87**	86%
Coral Rubble Colonized with Mixed Algae					1	**11**		**12**	92%
Live & Upright Dead Coral Colonized with Mixed Algae	2		2		1	2	**55**	**62**	89%
n_i	32	14	54	5	84	19	65	**273**	
Producer's Accuracy (%)	81%	79%	94%	100%	89%	58%	85%		

OA = 85.7%
Tau= 0.83
CI (±) = 0.06

3.5 CHANGES IN LAGOON HABITATS 2001-2016

Measuring habitat change by comparing field data from 252 AA sites visited in 2016 to the 2001 IKONOS mosaic revealed change at 11.5% of sites overall (Table 3.2). Fewest changes were observed on 'Seagrass (*E. acoroides*) on Sand' (0%) and 'Pavement Colonized with Mixed Algae' (3%) habitats. The greatest number of habitat changes observed were among the 'Mixed Algae and Seagrass on Sand' (26%) and 'Seagrass (*H. uninervis*) on Sand' (21%) categories. Of the 55 sites classified as 'Bare Sand' in the 2016 data, 7 (13%) were formerly occupied by seagrass in 2001. The 'Coral Rubble Colonized with Mixed Algae' (n=18), 'Seagrass (*H. uninervis*) on Sand' (n=14), and especially the 'Seagrass (*E. acoroides*) on Sand' (n=5) habitat classes, had very few sites, and are best interpreted in combination with the additional 50 randomized comparison points in the next section.

Results and Discussion

Measuring habitat change by comparing 350 sites in the 2016 WV2 image stratified by habitat class to the 2001 IKONOS mosaic revealed change at 16.6% of sites overall (Table 3.3). Changes among habitat types were similar to those observed using the AA data (Table 3.3). Fewest changes were observed on 'Seagrass (*E. acoroides*) on Sand' (12%) and 'Pavement' (10%) habitats, and the greatest changes were observed among the 'Mixed Algae and Seagrass on Sand' (30%) and 'Seagrass (*H. uninervis*) on Sand' (22%) categories. Of the 50 sites classified as 'Bare Sand' in the 2016 data, 8 (16%) were formerly occupied by seagrass in 2001.

The results of these two point-based assessments were similar despite differing inputs. One approach compared *in situ* field data collected in 2016 to visual interpretation of habitats from the 2001 IKONOS mosaic. The other approach compared model output from 2016 WV2 imagery to visual interpretation of the 2001 IKONOS mosaic. Taken together, these two approaches suggest that changes to the Lagoon habitats since 2001 were generally not widespread. The greatest number of changes were observed among 'Bare Sand', 'Seagrass (*H. uninervis*) on Sand', and 'Mixed Algae and Seagrass on Sand' habitats. In other words, the underlying substrate was not altered, only the density and proportional colonization of algae and *H. uninervis* fluctuated. Around 15% of 'Bare Sand' sites in 2016 appeared occupied by seagrass in 2001, suggesting moderate losses. Also of note, similar numbers of sites dominated with *H. uninervis* switched to mixed macroalgae and vice versa. A much smaller number of sites shifted among the hard bottom categories.

Although these two analyses did not suggest broad-scale changes in Lagoon habitats, they are based on the seven general habitat classes. The outcome could have been different if more detailed habitat classifications had been used. For example, several areas identified as living *Acropora* reefs in 2001 imagery (Houk and van Woesik, 2008) were observed to be largely dead, colonized by turf algae, and the few remaining corals were bleached, during 2016 field assessments. This represents a major change in these patch reef systems not quantified in this assessment due to the necessity of combining live and dead upright corals into one classification in the composite map. Despite the importance of live coral as a discrete habitat type and its attempted use in other recent mapping projects in the Lagoon (NOAA NCCOS, 2005; Houk and van Woesik, 2008), several factors prevented its use here. First, prior maps delineated large polygons wherein live coral may have been proportionally small but was prioritized in the attributes. The present maps are raster-based and do not prioritize one bottom type over another. In addition, as noted above with the loss of formerly monotypic patches of Staghorn *Acropora*, the extent of live coral in the Lagoon has changed markedly since the last maps were created. Where live coral did occur it was seldom the dominant cover (see GV and AA data), was mixed with other bottom types, and could not be strongly correlated with any of the 24 environmental predictors. Similarly, seagrass to macroalgae shifts over the last ten years have recently been investigated in detail using transects at 12 sites in the Lagoon (Camacho, 2016). Macroalgal canopies were associated with watershed size and development and appear to be increasing in seagrass beds around developed areas. The habitat classes used in the change analyses of the present study are not ideally suited to detect gradual shifts in macroalgal/seagrass proportions, but the raw videos and field data could serve as another point in the time series to monitor such change. Another important caveat is that (except for cloudy areas in the 2001 IKONOS mosaic) the scope of inference for these randomized, point-based approaches included nearly the entire Lagoon. However, any concentrated changes between assessment points were not quantified, and there were five areas that experienced dramatic changes in habitat from 2001-2016.

Changes at these five locations are qualitatively described with side-by-side satellite images from 2001 and 2016. This evaluation was supplemented with on-screen digitizing (1:1000) to estimate the size of the affected features in m^2 and in situ photos of the sites from 2016 where available. An important caveat in these examples is that only large areas of change are digitized for illustration purposes. Other changes are visible but were not quantified, and specific mechanisms for habitat changes are not defined. The five areas highlighted are: 1) north of the Sugar Dock, 2) west of Red Beach, 3) Garapan and Memorial Park, 4) Tanapag, and 5) northeast of Mañagaha.

Table 3.2. Changes at AA points in each of the seven habitat classes. Off diagonal cell values denote the number of habitat shifts from 2001 to 2016.

AA Sites \ 2001 (IKONOS)	Mixed Algae & Seagrass on Sand	Seagrass (*Halodule uninervis*) on Sand	Bare Sand	Seagrass (*Enhalus acoroides*) on Sand	Pavement Colonized with Mixed Algae	Coral Rubble Colonized with Mixed Algae	Live & Upright Dead Coral Colonized with Mixed Algae	n	Changed (%)
Mixed Algae & Seagrass on Sand	**20**	3	2	1			1	**27**	26%
Seagrass (*Halodule uninervis*) on Sand	1	**11**	1		1			**14**	21%
Bare Sand	4	3	**48**					**55**	13%
Seagrass (*Enhalus acoroides*) on Sand				**5**				**5**	0%
Pavement Colonized with Mixed Algae		2			**66**			**68**	3%
Coral Rubble Colonized with Mixed Algae			1			**15**	2	**18**	17%
Live & Upright Dead Coral Colonized with Mixed Algae	1		1		2	3	**58**	**65**	11%
								252	

Overall Change = 11.5%

Table 3.3. Changes at 50 random points in each of the seven habitat classes. Off diagonal cell values denote the number of habitat shifts from 2001 to 2016.

2016 Composite \ 2001 (IKONOS)	Mixed Algae & Seagrass on Sand	Seagrass (*Halodule uninervis*) on Sand	Bare Sand	Seagrass (*Enhalus acoroides*) on Sand	Pavement Colonized with Mixed Algae	Coral Rubble Colonized with Mixed Algae	Live & Upright Dead Coral Colonized with Mixed Algae	n	Changed (%)
Mixed Algae & Seagrass on Sand	**35**	14		1				**50**	30%
Seagrass (*Halodule uninervis*) on Sand	10	**39**	1					**50**	22%
Bare Sand	3	5	**42**					**50**	16%
Seagrass (*Enhalus acoroides*) on Sand	6			**44**				**50**	12%
Pavement Colonized with Mixed Algae		2			**45**		3	**50**	10%
Coral Rubble Colonized with Mixed Algae	1	1	1		1	**43**	3	**50**	14%
Live & Upright Dead Coral Colonized with Mixed Algae	1				4	1	**44**	**50**	12%
								350	

Overall Change = 16.6%

Results and Discussion

Sugar Dock

The seagrass meadow north of the Sugar Dock has undergone noticeable change from 2001 to 2016 (Figure 3.15a). In 2001, the area was dominated by an extensive *H. uninervis* meadow with continuous coverage and high shoot density (Figure 3.15b). By 2016, a large part of this meadow had been converted to a patchy seagrass habitat (Figure 3.15c). Digitizing of the area suggested that approximately 90,000 m^2 changed from dense to patchy seagrass and appears to be associated with a shift of 50-100 m shoreward of increased sand coverage and/or reduced seagrass density. Field images collected in 2016 confirm that this site remains colonized with *H. uninervis* however, patches of bare sand are now evident (Figure 3.15d).

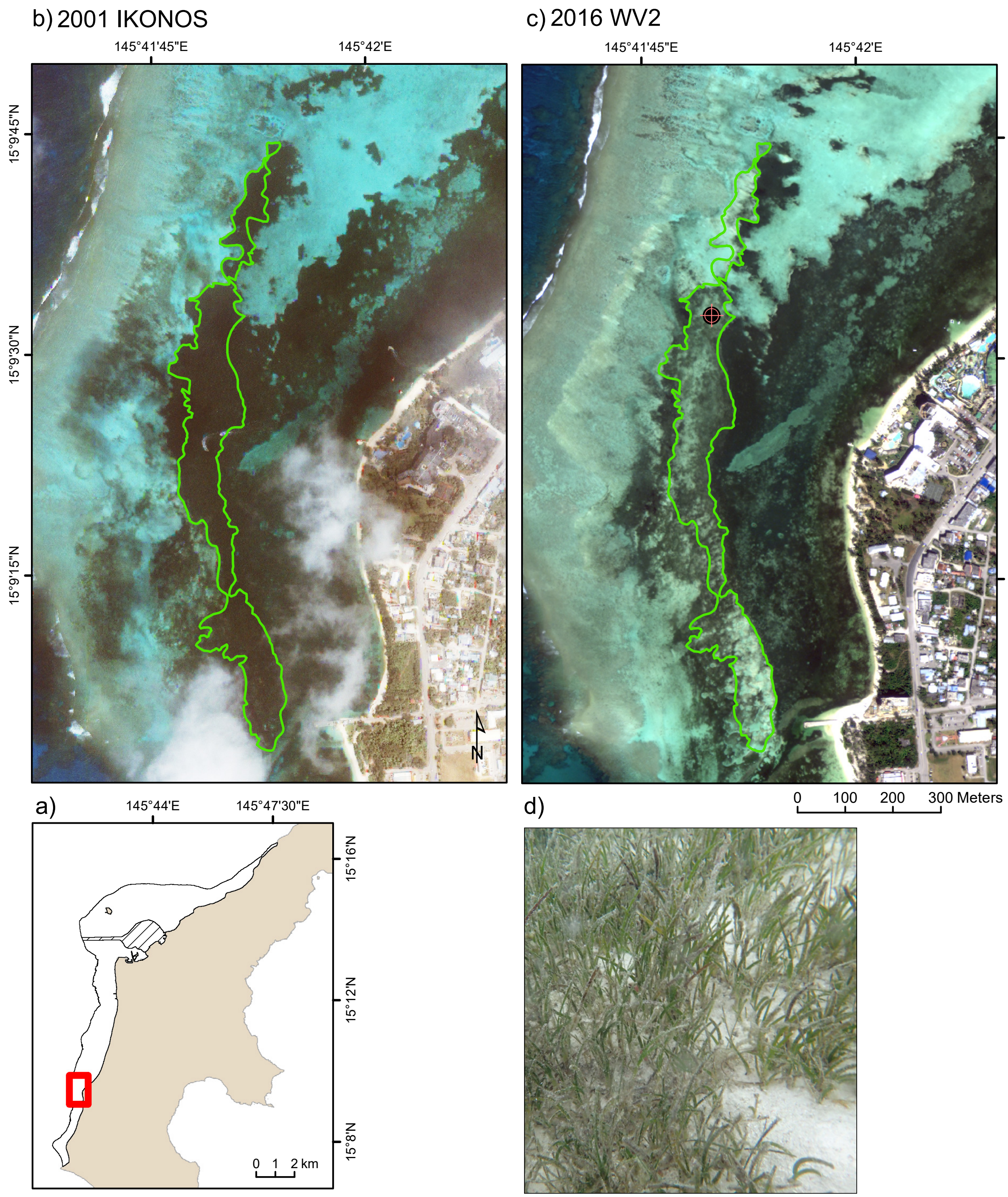


Figure 3.15. Habitat changes at Sugar Dock. Clockwise from lower left: a) location of the change within the Lagoon, b) 2001 IKONOS mosaic of the site with the green polygon denoting the specific feature of change, c) 2016 WV2 image of the site with the same polygon denoting the area of change, and d) in situ *photo from the 2016 field data (location denoted by the crosshairs symbol on 3.15c).*

Red Beach

Lagoon habitats, including seagrass beds and Staghorn *Acropora* reefs off Red Beach, have undergone noticeable change from 2001 to 2016 (Figure 3.16a). In 2001, there was a large *H. uninervis* meadow with continuous coverage and high shoot density located approximately mid-Lagoon (green outline in Figure 3.16b). By 2016, the entire southern 2/3 (approximately 40,000 m^2) of this meadow had been converted to a sand and patchy seagrass habitat (Figure 3.16c). Sand deposition may be the most likely cause of the loss. Additional changes to small *H. uninervis* patches are also evident to the east of this large patch. North of these seagrass are several dark features oriented East-West. Field observations confirm that these were Staghorn *Acropora* reefs. These have experienced changes in shape and patchiness from 2001-2016, including the growth of approximately 10,000 m^2 of additional reef in the example shown here (red outline in Figure 3.16c). Unfortunately, field images collected in August 2016 at these and other similar reefs, indicate that most of the Staghorn is dead, colonized with turf algae, and remaining live corals were bleached (Figure 3.16d). It is unknown whether these corals recovered following the bleaching event.

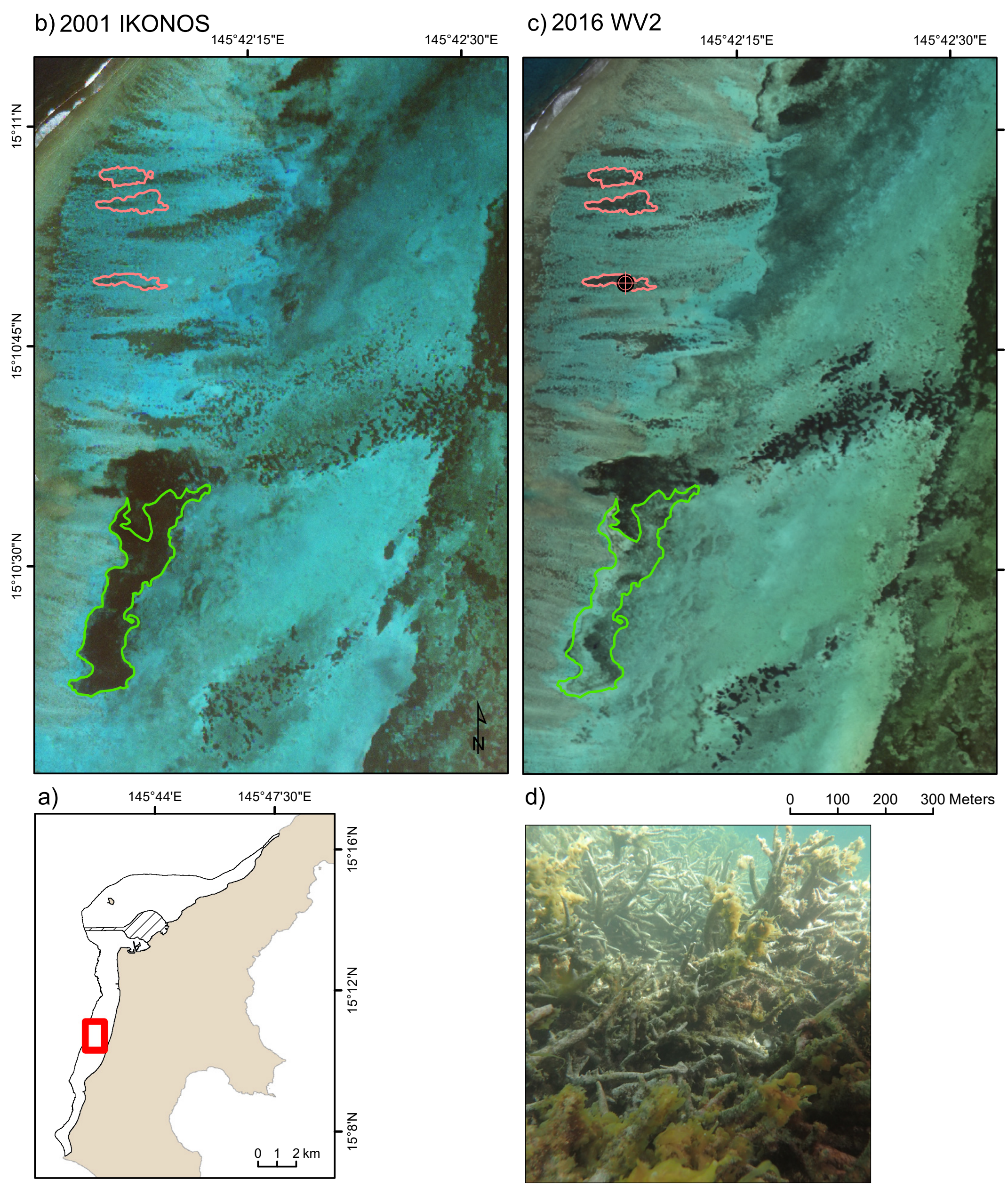


Figure 3.16. Habitat changes at Red Beach. Clockwise from lower left: a) location of the change within the Lagoon, b) 2001 IKONOS mosaic of the site with the green and red polygons denoting the specific features of change, c) 2016 WV2 image of the site with the same polygons denoting the areas of change, and d) in situ *photo from the 2016 field data (location denoted by crosshairs symbol on 3.16c).*

Garapan/Memorial Park

There are several examples of habitat change in the Garapan and Memorial Park area from 2001 to 2016 (Figure 3.17a). Since 2001, the western beach at Memorial Park has receded 40-50 m whereas the beach off the Hyatt Regency hotel has grown seaward 20-30 m (orange outline based on 2001 in Figures 3.17a-b). Field observations in 2016 documented the erosion along the point off Memorial Park. Looking southward in the photograph where there is presently shallow water and broad sandy beach (Figure 3.17d), was previously coastal forest in 2001. Extensive sand movement is also evident in this area as it related to changes in extent of seagrass meadows. Beach profile studies of the shoreline at Memorial Park and southward through the hotel zone to Susupe provide a detailed documentation of shoreline movement along this coast prior to Soudelor (USACE, 2004; Yuknavage and Palmer, 2010). Three large *H. uninervis* meadows with relatively continuous coverage and high shoot density in 2001 (green outline in Figure 3.17b) have now been converted to a sand and patchy seagrass habitat or even pavement in 2016 (Figure 3.17c). Scour marks are evident in several spots in the 2016 imagery, where the Lagoon floor has been stripped down to bare pavement.

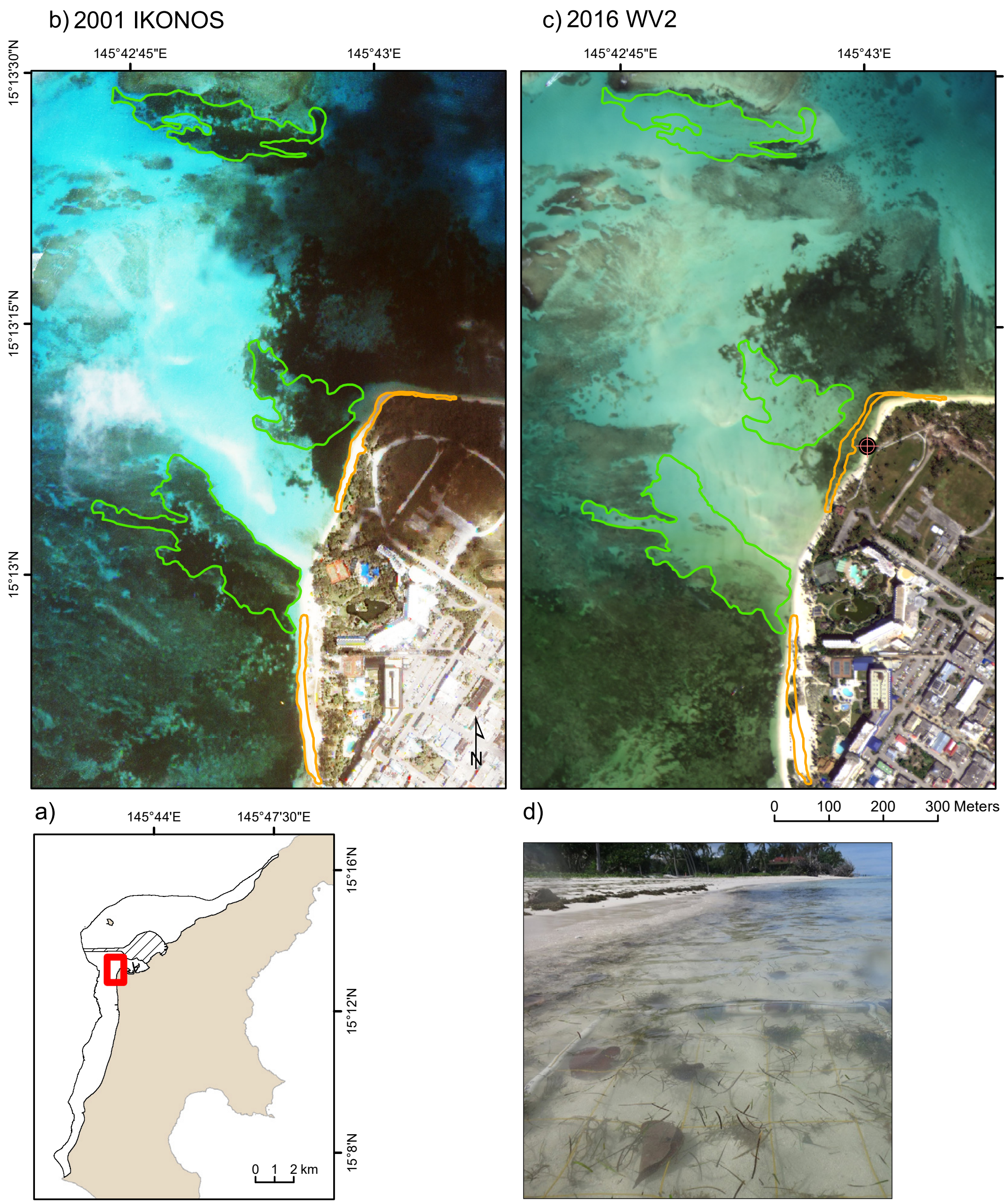


Figure 3.17. Habitat changes at Garapan/Memorial Park. Clockwise from lower left: a) location of the change within the Lagoon, b) 2001 IKONOS mosaic of the site with the green and orange polygons denoting the specific features of change, c) 2016 WV2 image of the site with the same polygons denoting the areas of change, and d) in situ *photo from the 2016 field data (location denoted by crosshairs symbol on 3.17c).*

Tanapag

The nearshore seagrass meadows off Tanapag have also been a notable area of change since 2001 (Figure 3.18a). The small point south of the boat ramp has accreted approximately 20 m of sand (orange outline in Figures 3.18b-c). More dramatically, the *H. uninervis* meadows in the area have lost approximately 210,000 m^2 of coverage since 2001 (green outlines in Figure 3.18c). Field images from 2016 suggest that although some areas have lost the dense and continuous coverage of *H. uninervis* that was present in 2001 and have been converted completely to sand, other areas still have sparse or patchy *H. uninervis* present (Figure 3.18d).

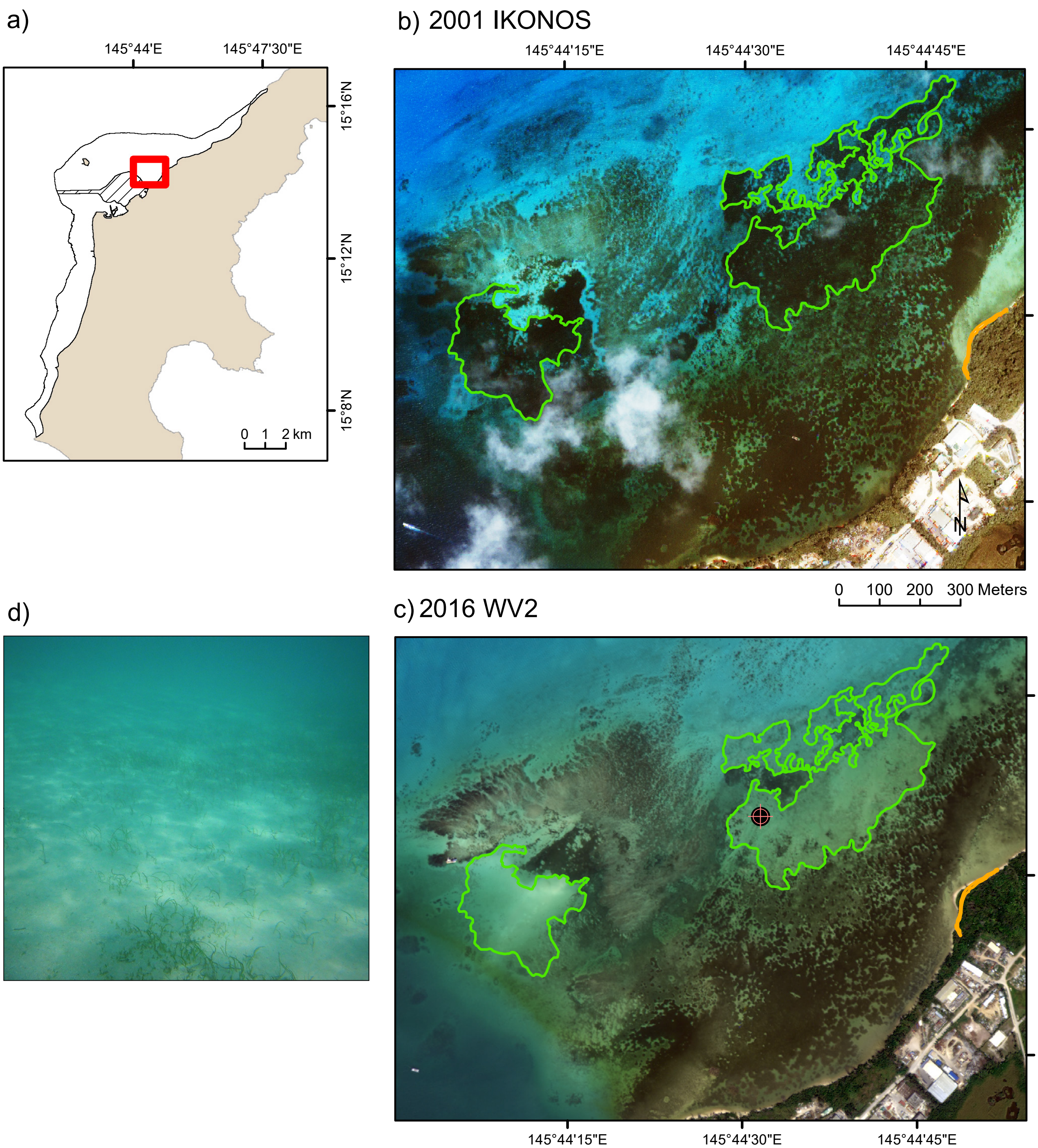


Figure 3.18. Habitat changes at Tanapag. Clockwise from upper left: a) location of the change within the Lagoon, b) 2001 IKONOS mosaic of the site with the green and orange polygons denoting the specific features of change, c) 2016 WV2 image of the site with the same polygons denoting the areas of change, and d) in situ *photo from the 2016 field data (location denoted by crosshairs symbol on 3.18c).*

Results and Discussion

Mañagaha

The popular tourist destination, Mañagaha Island, has also undergone two notable changes in its surrounding habitats between 2001 and 2016 (Figure 3.19a). First, of primary concern to tourism, the sand beach on the southeastern shoreline has continued its erosion since 2001, but has accreted 20-30 m in places north and west of the island (orange line in Figure 3.19b-c; see Fletcher et al., 2007 for detailed information on shoreline change and its causes at Mañagaha up to 2006). Also of note, the patch reefs to the northeast of the island appear to have grown considerably since 2001 (red lines in Figure 3.19b-c). Field data in 2016 indicates that these are comprised primarily of Staghorn *Acropora* (Figure 3.19d) and *Isopora* colonies, and cover an area of approximately 115,000 m^2. As was observed in many places, much of the Staghorn coral was dead and covered with turf algae but still upright. The remainder of live Staghorn was bleached at the time of this field image in August 2016.

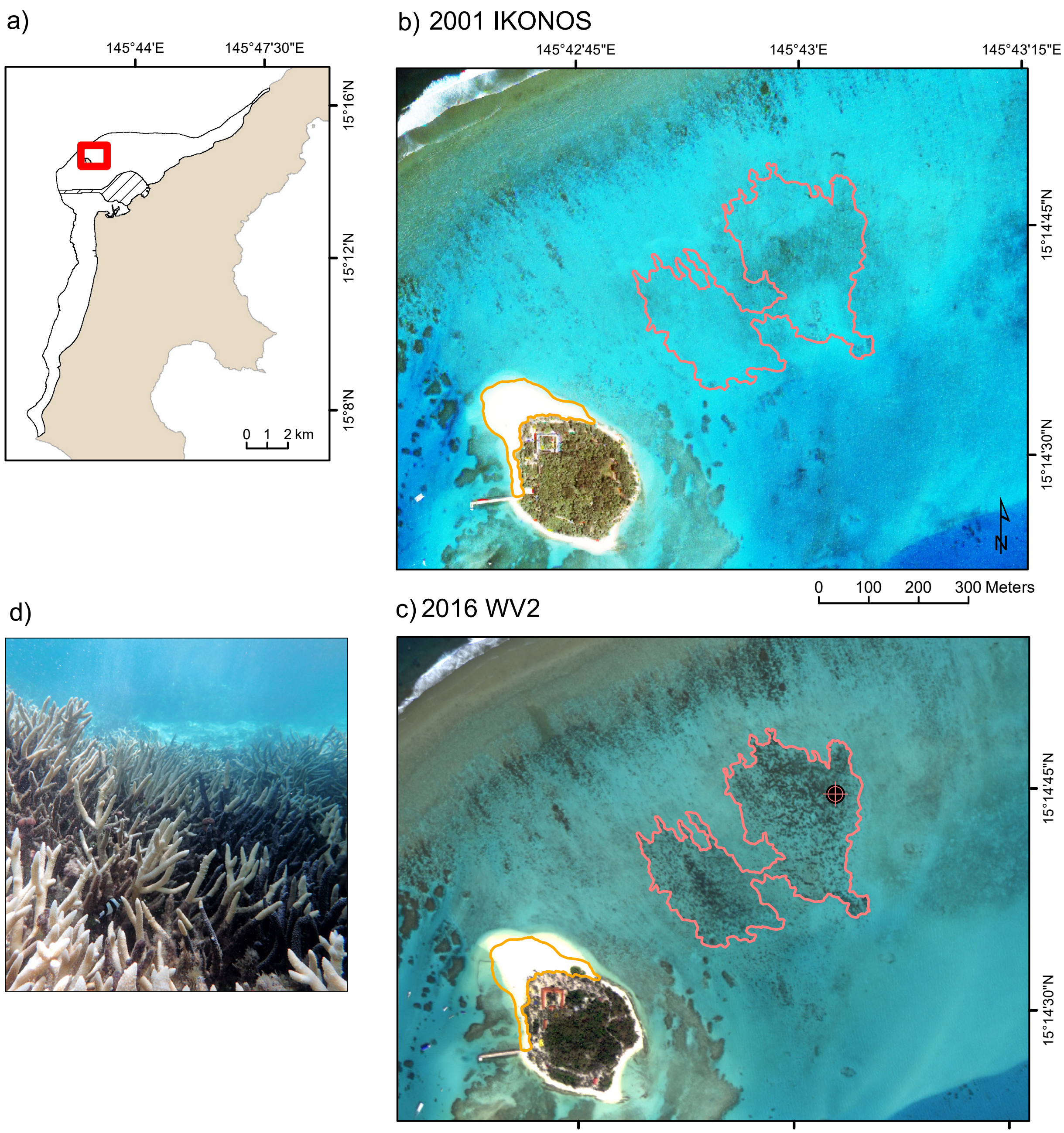


Figure 3.19. Habitat changes at Mañagaha. Clockwise from upper left: a) location of the change within the Lagoon, b) 2001 IKONOS mosaic of the site with the red and orange polygons denoting the specific features of change, c) 2016 WV2 image of the site with the same polygons denoting the areas of change, and d) in situ *photo from the 2016 field data (location denoted by crosshairs symbol on 3.19c).*

3.6 APPLICATIONS

The first applications of the composite map are anticipated to be for the SLUMP update underway in 2017 and the revised economic valuation in 2018 (van Beukering et al., 2006; SLUMP, 2012). Habitat maps can be used to plan for potential Lagoon activities in a way that will reduce conflict and minimize impacts to important habitats. They are also needed for identifying different bottom features and estimating their economic value per-unit-area. Habitat maps are also ideal for planning other spatial-management decisions on topics such as identifying and quantifying essential fish habitat, calculating damage and costs following ship grounding or other impacts, monitoring habitat or shoreline changes through time, minimizing development impacts to important habitats, sampling design for monitoring or scientific studies, educational materials, and planning for marine protected areas or other areas of particular concern (USACE, 2004; Fletcher et al., 2007; CNMI and NOAA CRCP, 2010; Yuknavage and Palmer, 2010; Costa et al., 2013; Kågesten et al., 2015).

There are a wide range of additional potential applications for the maps, underwater videos, and other datasets associated with this project. The best way to access and use these highly detailed maps will be through GIS or other software that allows users to zoom in to a custom scale to suit their particular purpose. We also recognize that at times, some users can better utilize a hard copy map displayed at reasonable scale and we therefore provide the composite habitat maps in an atlas format at the end of this report (Appendix A). The atlas consists of six color maps with corresponding WV2 satellite imagery at a printed scale of 1:17,000 such that the 2 x 2 m prediction grid can be discerned.

Some local users familiar with the earlier polygon-based maps (4,047 m^2 MMU as in NOAA NCCOS, 2005; Houk and van Woesik, 2008) may initially be challenged by the transition to these new, highly resolved (2 x 2 m) raster-based maps. Spatial and spectral resolution of satellite sensors, computing power, and model-based mapping techniques have advanced considerably since the last time Saipan Lagoon's benthic habitats were characterized. The map products created here take maximum advantage of those improvements and preserve the fine-scale heterogeneity, habitat gradients, and smaller features present in the real landscape. Reducing the resolution to match previous maps may be useful, but the process should be customized to suite specific applications. Therefore, the full potential and spatial detail of the basic products was not diminished. All of the potential modifications discussed below can be implemented using the basic tools in GIS or other software that can manipulate grid data.

Some options for customization may include use of spatial filters to enhance dominant or priority bottom types, or to smooth out variability in heterogeneous areas. Habitat classes can also be aggregated to reduce thematic nuances (e.g., hard bottom versus soft bottom rather than multiple types of reef and sand habitats) or translated into other classification systems (e.g., CMECS, 2017). Predictions for individual substrate and cover types can be converted from the continuous probability-of-occurrence format that is good for examining gradients, into specific categories or cut-offs. For example, transitions from bare sand through algae or seagrass cover can be adjusted into discrete categories to better characterize subtle or more dramatic shifts in presence.

In addition to the digital products accessible through GIS software, we also provide an on-line data viewer that allows users without any specialized software to view and query the WV2 satellite image, habitat maps, field data including high definition underwater videos, as well as model predictor data. The underwater videos can also be downloaded directly from the data viewer. Access these products through our website at: https://coastalscience.noaa.gov/projects/detail?key=271

Bleached and recently dead Staghorn coral (Acropora species). Credit: M. Kendall and B. Costa.

References

Battista, T.A., B.M. Costa, and S.M. Anderson. 2007. Shallow-Water Benthic Habitats of the Republic of Palau. NOAA Technical Memorandum. NOS NCCOS 59. Silver Spring, MD.

Bindel, M., T. Heege, K. Hartmann, and H. Needham. 2013. Final report satellite derived bathymetry map Saipan and Tinian, Northern Mariana Islands. Task Order S/C EA133C-11-CQ-0007-001. 13 pp.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.I. Stone. 1984. Classification and regression trees. Taylor & Francis, Belmont, CA. 368 pp.

Camacho, R.C. 2016. Drivers of macroalgal canopy dynamics within Saipan Lagoon seagrass habitats. Master of Science Thesis. University of Guam. 52 pp.

Card, D.H. 1982. Using known map categorical marginal frequencies to improve estimates of thematic map accuracy. Photogrammetric Engineering and Remote Sensing 48: 431-439.

Cloud, P.E. 1959. Geology of Saipan Mariana Islands, Part 4: Submarine Topography and Shoal-water Ecology. Geological Survey Professional Paper 280-K. Washington, DC. USA

CMECS. 2017. Coastal and Marine Ecological Classification Standard, Catalog of Units. Online: https://www.cmecscatalog.org/ (Site Accessed 16 February 2017)

CNMI and CRCP (Commonwealth of the Northern Mariana Islands and NOAA Coral Reef Conservation Program). 2010. Commonwealth Of The Northern Mariana Islands' Coral Reef Management Priorities. Commonwealth of the Northern Mariana Islands and NOAA Coral Reef Conservation Program. Silver Spring, MD. 28 pp.

Congalton, R.G., and K. Green. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC/Lewis Press, Boca Raton, FL. 137 pp.

Costa, B.M., M.S. Kendall, K. Edwards, G. Kågesten, and T.A. Battista. 2013. Benthic Habitats of Fish Bay, Coral Bay and the St. Thomas East End Reserve. NOAA Technical Memorandum NOS NCCOS 175. Silver Spring, MD. 68 pp.

De'ath, G., and K.E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.

De'ath, G. 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243-251.

Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J.M. Overton, A. Townsend Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

Elith J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77:802-81.

ESRI. 2016. ArcGIS (Version 10.4). Environmental Systems Research Institute, Redlands, California. Software Online: http://www.esri.com/ (Site Accessed 22 February, 2016)

References

Fletcher, C.H., M. Barbee, M. Dyer, A. Genz, and S. Vitousek. 2007. Mañagaha Island Shoreline Stability Assessment. Report to the Coastal Resources Management Office, CNMI, Saipan. 90 pp.

Friedman, J.H. 2002. Stochastic gradient boosting. Computational Statistics and Data Analysis 38:367-378.

Guard, C., and M.A. Lander. 2015. Meteorological assessment for Typhoon Soudelor in Saipan, CNMI. National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) Meteorological Assessment Team for Typhoon Soudelor. NWS Guam and University of Guam. 49 pp.

Hijmans, R.J. 2014. R package, Raster: Geographic data analysis and modeling. Software Downloaded October 2014. Software Website: http://CRAN.R-project.org/package=raster (Site Accessed 8 June 2016).

Hijmans, R.J., S. Phillips, J. Leathwick, and J. Elith. 2014. R package, dismo: Species distribution modeling. Software Downloaded October 2014. Software website: http://CRAN.R-project.org/package=dismo (Site Accessed 8 June 2016).

Hosmer, D.W., and S. Lemeshow. 2000. Applied logistic regression (2nd ed.). John Wiley & Sons. 375 pp.

Houk, P., and R. van Woesik. 2008. Dynamics of shallow-water assemblages in the Saipan Lagoon. Marine Ecology Progress Series 356:39-50.

Jenness, J. 2016. DEM Surface Tools (Version 2.1.399). Jenness Enterprises, Flagstaff, AZ. Software downloaded October 2015. Software Website: http://www.jennessent.com/arcgis/surface_area.htm (Site Accessed 22 February, 2016)

Kågesten, G., W. Sautter, K. Edwards, B. Costa, L. Kracker, and T. Battista. 2015. Shallow-Water Benthic Habitats of Northeast Puerto Rico and Culebra Island. NOAA Technical Memorandum NOS NCCOS 200. Silver Spring, MD. 111 pp.

Kendall, M.S., and T. Miller. 2008. The influence of thematic and spatial resolution on maps of a coral reef ecosystem. Marine Geodesy 31:75-102.

Kuhn, M. 2016. R package, Caret: Classification and Regression Training. Software Downloaded October 2014. Software Website: https://cran.r-project.org/web/packages/caret/index.html (Site Accessed 22 February 2017).

Leathwick, J.R., J. Elith, M.P. Francis, T. Hastie, and P. Taylor. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series 321:267-281.

Lyzenga, D.R. 1978. Passive remote sensing techniques for mapping water depth and bottom features. Applied Optics, 17:379-383.

Ma, Z., and R.L. Redmond. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. Photogrammetric Engineering and Remote Sensing 61:435-439.

Mariana Visitors Authority. 2017. Commonwealth of the Northern Mariana: Visitor Arrival Statistics FY 2016.

(Accessed 2 February 2017).

MGET (Marine Geospatial Modeling Tools). 2016. Marine Geospatial Ecology Tools: Open source geoprocessing for marine research and conservation (version 0.8a64). Duke Nicholas School of the Environment, Marine Geospatial Ecology Lab. Software downloaded June 2016. Software Online: http://mgel2011-kvm.env.duke.edu/mget/download/ (Site Accessed 22 February, 2017).

Mumby, P.J., C.D. Clark, E.P. Green, and A.J. Edwards. 1998. Benefits of water column correction and contextual editing for mapping coral reefs. International Journal of Remote Sensing 19:203-210.

Mumby, P.J., and A. Edwards. 2000. Chapter 8: Water Column Correction Techniques. In: E.P. Green, P.J. Mumby, A.J. Edwards and C.D. Clark, Remote Sensing Handbook for Tropical Coastal Management. UNESCO Publishing: Paris, France. 316 pp. Online: http://www.unesco.org/csi/pub/source/rs.htm (Accessed 21 February 2017).

NOAA. 2002. A national coral reef action strategy. Report to congress on implementation of the Coral Reef Conservation Act of 2000 and the National Action Plan to Conserve Coral Reefs in 2002-2003. Silver Spring, MD. 156 pp.

NOAA COOPS (Center for Operational Oceanographic Products and Services). 2016. Data downloaded 8 September 2016. Tides & Currents, Tides/Water Levels, 1630000 Apra Harbor, Guam. Data website: https://tidesandcurrents.noaa.gov/waterlevels.html?id=1630000&units=metric&bdate=20160721&edate=20160801&timezone=GMT&datum=MLLW&interval=6&action=data (Accessed 22 February 2017).

NOAA NCCOS (National Centers for Coastal Ocean Science). 2001. Atlas of the Shallow-Water Benthic Habitats of American Samoa, Guam, and the Commonwealth of the Northern Mariana Islands; CNMI – Data; Saipan; Image Mosaics(ZIP, 41.2KB). Data downloaded 5 March 2016. Data Website: https://products.coastalscience.noaa.gov/collections/benthic/e99us_pac/data_cnmi.aspx (Site Accessed 21 February 2017).

NOAA NCCOS (National Centers for Coastal Ocean Science). 2002. Atlas of the Shallow-Water Benthic Habitats of American Samoa, Guam, and the Commonwealth of the Northern Mariana Islands; CNMI – Data; Saipan; GPS Control Points (ZIP, 11KB). Data downloaded 5 March 2016. Data Website: https://products.coastalscience.noaa.gov/collections/benthic/e99us_pac/data_cnmi.aspx (Site Accessed 21 February 2017).

NOAA NCCOS (National Centers for Coastal Ocean Science). 2005. Atlas of the Shallow-Water Benthic Habitats of American Samoa, Guam, and the Commonwealth of the Northern Mariana Islands. NOAA Technical Memorandum NOS NCCOS 8. Silver Spring, MD. 126 pp. Online: https://products.coastalscience.noaa.gov/collections/benthic/e99us_pac/ (Accessed 16 February 2017).

Northern Mariana Islands Tourism Master Plan. 2012. Northern Mariana Islands Tourism Master Plan 2012-2016. Prepared for the Marianas Visitors Authority by Management Analysis, Inc. Vienna, VA. 181 pp.

SLUMP. 2012. Saipan Lagoon Use Management Plan Update. Prepared for the Commonwealth of the Northern Mariana Islands Coastal Resources Management Office by Tetra Tech, Honolulu, HI. 104 pp.

Story, M., and R. Congalton. 1986. Accuracy assessment: A user's perspective. Photogrammetric Engineering and Remote Sensing 52:397-399.

R Core Team. 2016. R: A language and environment for statistical computing (Version 3.3.0). R Foundation

References

for Statistical Computing, Vienna, Austria. Software downloaded September 2015. Software Online: https://www.r-project.org/ (Site Accessed 22 February, 2016).

USACE (U.S. Army Corps of Engineers). 2004. Saipan Lagoon Erosion Study: Saipan Island, Commonwealth of the Northern Mariana Islands. Honolulu Engineer District. Fort Shafter, HI. 56 pp.

USACE (U.S. Army Corps of Engineers). 2007. 2007 USACE Lidar: Pacific Islands (Saipan). Data downloaded 5 March, 2016. Data Website: https://coast.noaa.gov/dataviewer/#/lidar/search/where:ID=561/details/561 (Accessed 21 February, 2016).

van Beukering, P. (editor). 2006. The Economic Value of the Coral Reefs of Saipan, Commonwealth of the Northern Mariana Islands. Prepared by: Cesar Environmental Economics Consulting. Arnhem, Netherlands. 163 pp.

Yuknavage, K. and A. Palmer. 2010. Monitoring and preparing for shoreline change at American Memorial Park (2004-2009). A report for the National Park Service. 21 pp.

Modelling Glossary

Note: many of these definitions are specific to the context of this project.

Bag fraction (bf) – In a boosting context, a parameter that defines the fraction of the data drawn at random, without replacement, from the full training dataset at each iteration.

Boosted classification tree model – A modeling approach that combines a machine learning technique, boosting, with traditional tree-based statistical modeling. In this approach, a large number of classification trees are fit stagewise (i.e., after each tree is fit, the remaining variation in the data is used to fit the next tree) and then combined to generate a final, ensemble model.

Boosted regression tree model – A modeling approach that combines a machine learning technique, boosting, with traditional tree-based statistical modeling. In this approach, a large number of regression trees are fit stagewise (i.e., after each tree is fit, the remaining variation in the data is used to fit the next tree) and then combined to generate a final, ensemble model.

Boosting – A technique for fitting models that employs an iterative approach. Models are built in a stage-wise fashion, where existing trees are left unchanged and the variance remaining from the last tree is used to fit the next one.

Bootstrapping – A data re-sampling technique for estimating the statistical precision (i.e., coefficient of variation) in model predictions.

Coefficient of variation (CV) – Measure of dispersion for a distribution, representing the standard deviation as a proportion of the mean. In the context of a model prediction, a larger CV indicates more variation (uncertainty) in the prediction relative to the mean prediction.

Environmental predictor – An independent variable in a model that is used to explain variation in the response.

K-fold cross-validation (kCV) – A technique for evaluating the predictive ability of a fitted model. The data are divided into k=10 data subsets (i.e., folds). Nine of these folds are used to create models, while the one remaining is used to evaluate the model's performance.

K-fold cross-validation percent deviance explained (kCV PDE) – Percent deviance explained calculated using the kCV test data and model predictions of the test data.

Learning rate (lr) – In a boosting context, the degree to which each tree contributes to the final model. The optimal learning rate is one that minimizes prediction error in the fewest number of boosting iterations.

Percent deviance explained (PDE) – A measure of the variation in the data explained by a model (beyond that explained by a model without predictor variables). Values normally range between 0 and 100%, although negative values are possible. Higher values indicate better model performance.

Receiver operating characteristic (ROC) Area under the curve (AUC) – An ROC curve is a graphical representation of how well a model can discriminate between (or predict) two categories of data (e.g., presence/absence). The AUC is the integral of a ROC curve. AUC values range between 0 and 1 where a value >0.5 indicates performance better than a random guess. Higher AUC values indicate better model performance.

Modelling Glossary

Resampling – A method of using randomly drawn subsets of data to estimate statistical precision (e.g., variation in model predictions), to perform a significance test (e.g., permutation test of predictor importance), or to perform model validation (e.g., cross-validation).

Root Mean Square Error (RMSE) – RMSE measures the error associated with a model by calculating the difference between the predicted values (extracted from the model) and the observed values (extracted from the underwater videos).

Spatial predictive modeling – Modeling technique whereby relationships between environmental predictors and a response variable are estimated for areas with field data. These relationships are then used to predict the response in areas without field data.

Sensitivity – Also known as the true positive rate, a measure of model performance for binary classification models (e.g., presences versus absence) that measures the proportion of positives that are correctly identified as positives. In the context of a MaxEnt model of habitat suitability, this is calculated as the fraction of presences correctly classified as suitable habitat.

Specificity – Also known as the true negative rate, a measure of model performance for binary classification models (e.g., suitable vs. unsuitable habitat) that measures the proportion of negatives that are correctly identified as negatives.

Test data – Data that are excluded during model fitting and later used to test the predictive performance of the fitted model (e.g., during cross-validation).

Test percent deviance explained (PDE) – PDE calculated for a fitted model using accuracy assessment (AA) test data.

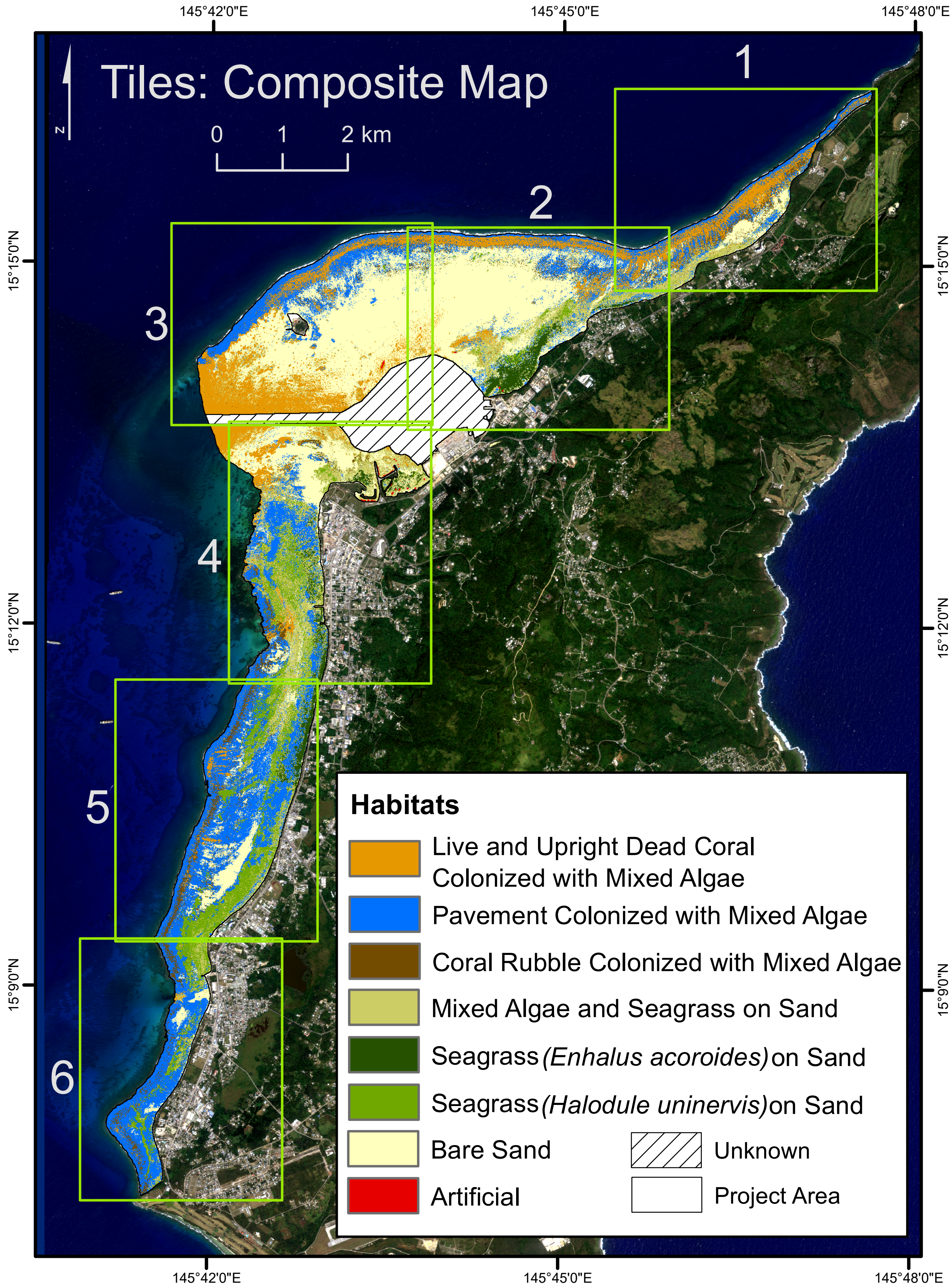
Training data – Ground validation (GV) data to which a model is fitted in order to estimate model parameter values.

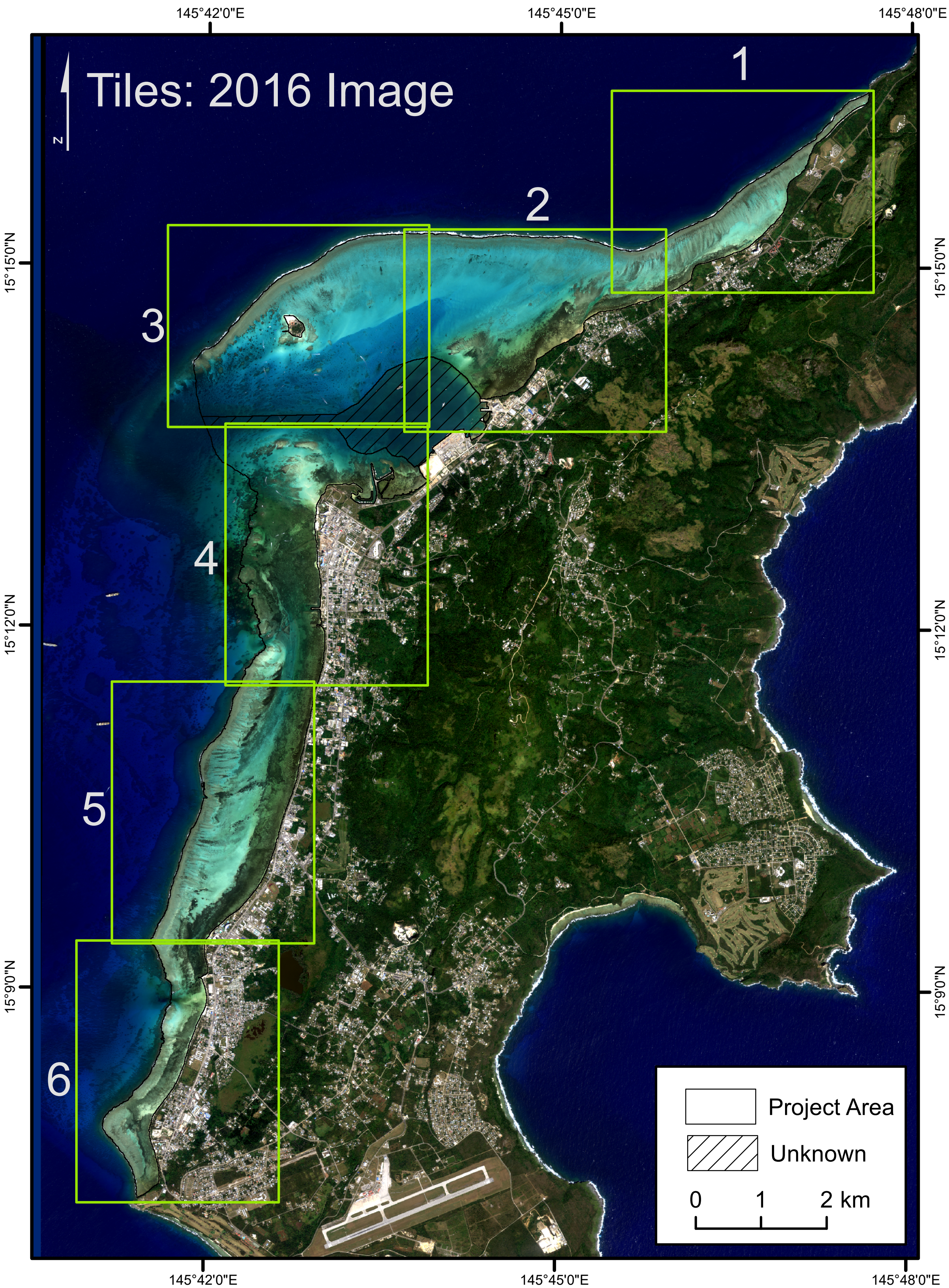
Tree complexity (tc) – In boosted regression and classification tree models, a parameter that controls the number of allowable nodes in a tree. This limits the number of possible interactions between predictor variables.

World War II Japanese airplane. Credit: M. Kendall and B. Costa.

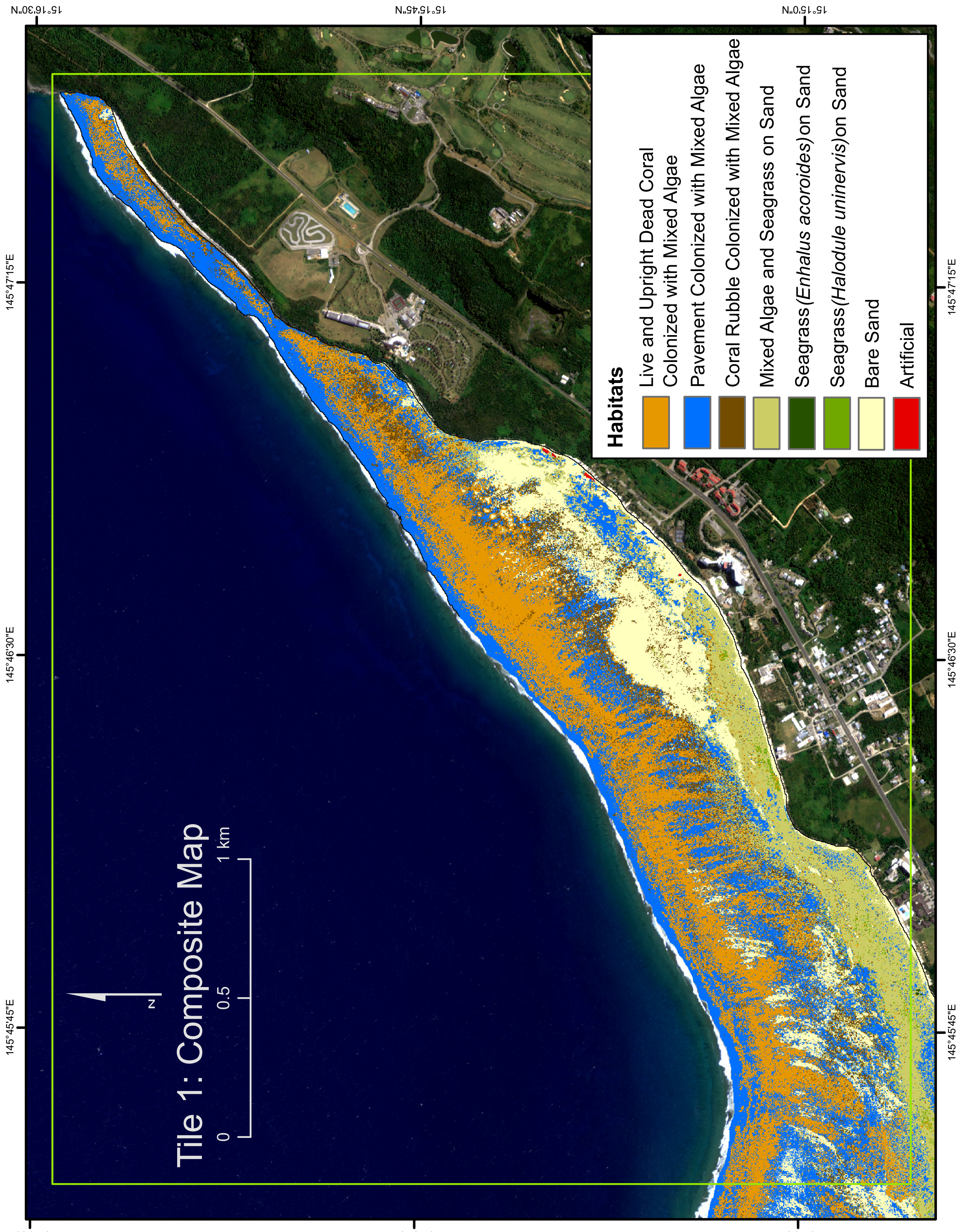
Appendix

Appendix - Tiles: These map tiles show more of the detail seen in the 2016 WV2 satellite image and benthic habitats for Saipan Lagoon. Please see the GIS datasets for the full resolution data.

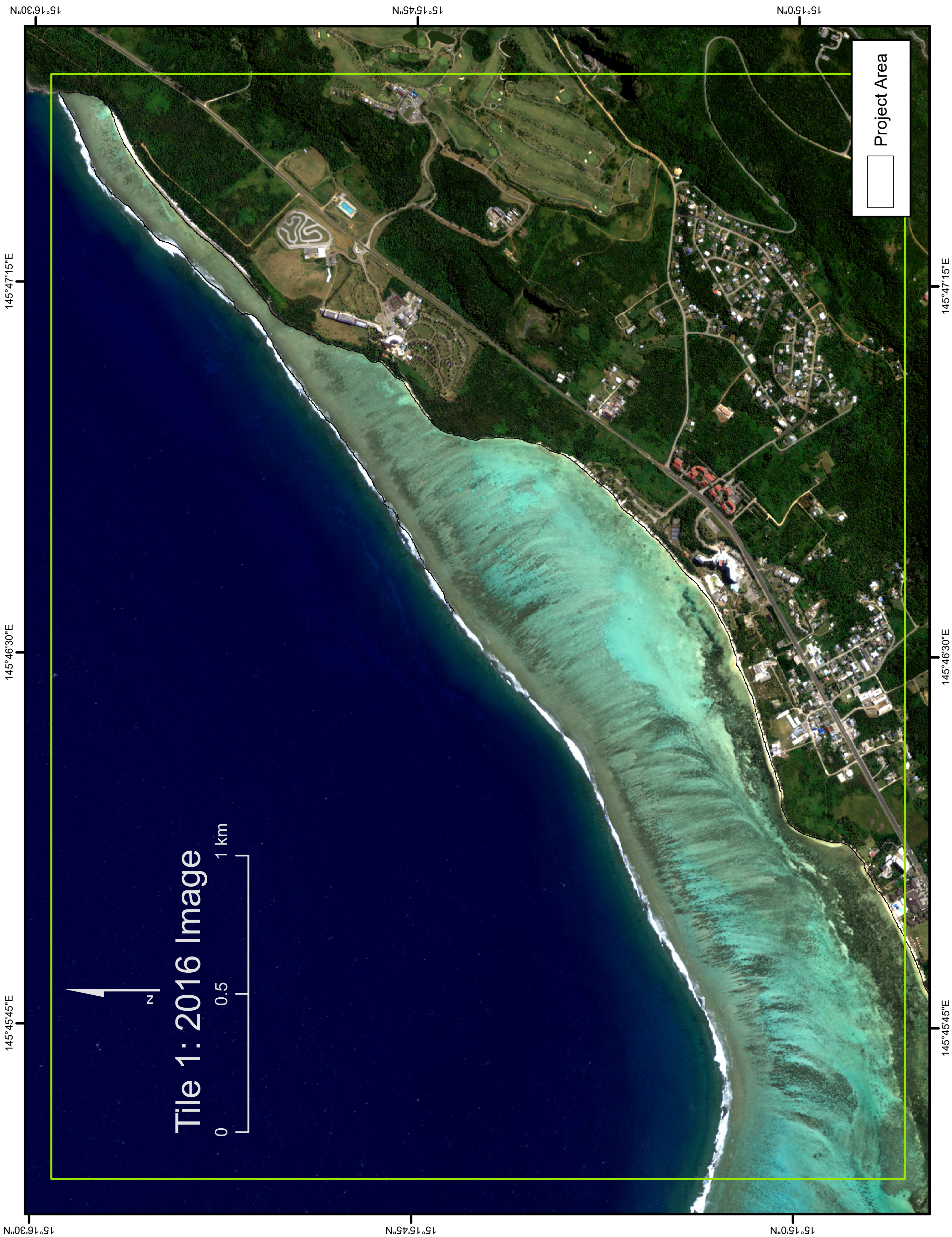



Tiles: 2016 Image
N
1
2
3
4
5
6
145°42'0"E
145°45'0"E
145°48'0"E
15°15'0"N
15°12'0"N
15°9'0"N
Project Area
Unknown
0 1 2 km

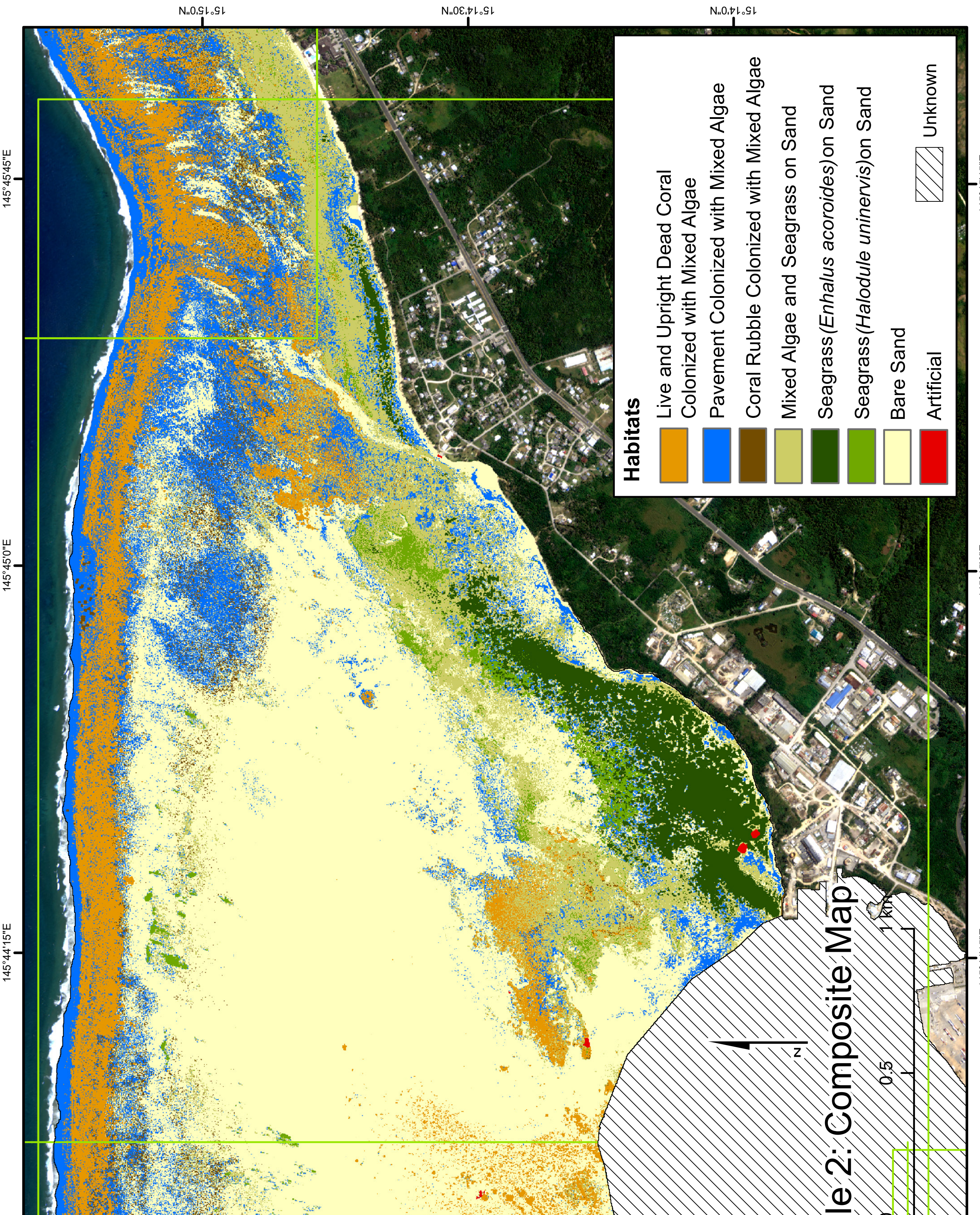
Appendix



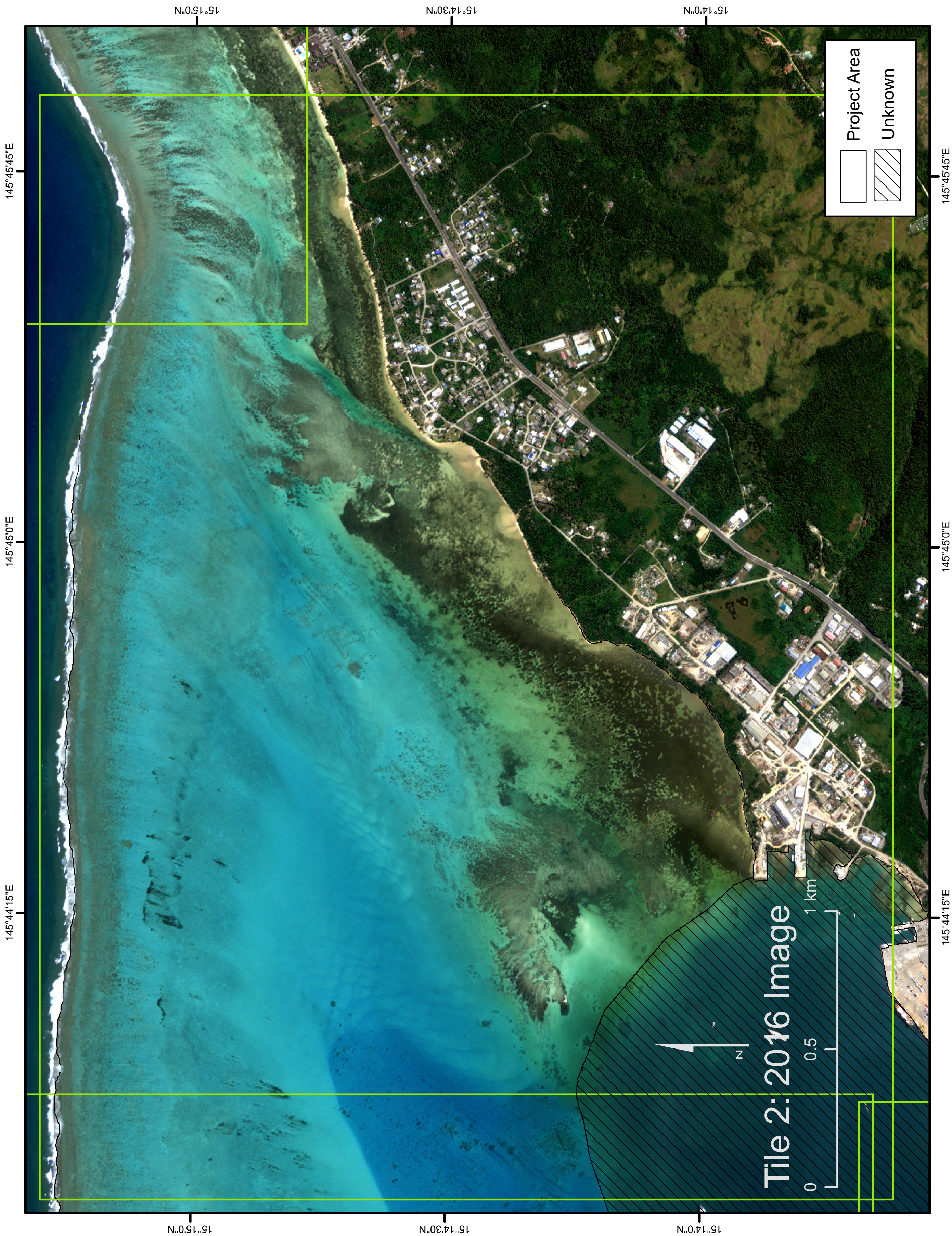
Tile 1: Composite Map
N
0
0.5
1 km
Habitats
Live and Upright Dead Coral Colonized with Mixed Algae
Pavement Colonized with Mixed Algae
Coral Rubble Colonized with Mixed Algae
Mixed Algae and Seagrass on Sand
Seagrass(*Enhalus acoroides*)on Sand
Seagrass(*Halodule uninervis*)on Sand
Bare Sand
Artificial
15°16'30"N
15°15'45"N
15°15'0"N
145°47'15"E
145°46'30"E
145°45'45"E


Tile 1: 2016 Image
Project Area
N
0
0.5
1 km
15°16'30"N
15°15'45"N
15°15'0"N
145°47'15"E
145°46'30"E
145°45'45"E

Appendix

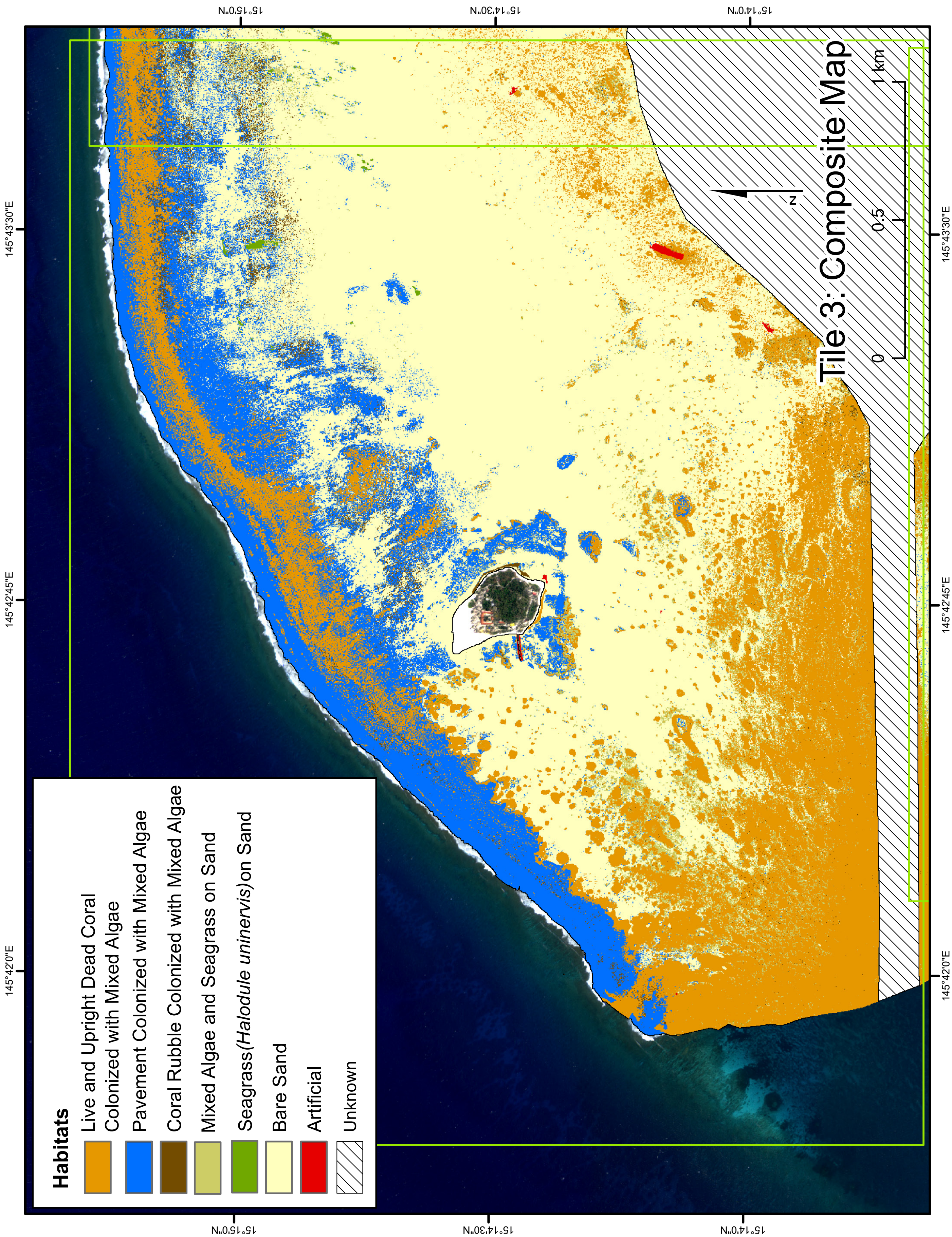


Tile 2: Composite Map
Habitats
Live and Upright Dead Coral Colonized with Mixed Algae
Pavement Colonized with Mixed Algae
Coral Rubble Colonized with Mixed Algae
Mixed Algae and Seagrass on Sand
Seagrass (*Enhalus acoroides*) on Sand
Seagrass (*Halodule uninervis*) on Sand
Bare Sand
Artificial
Unknown
N
0
0.5
1 km
15°15'0"N
15°14'30"N
15°14'0"N
145°45'45"E
145°45'0"E
145°44'15"E

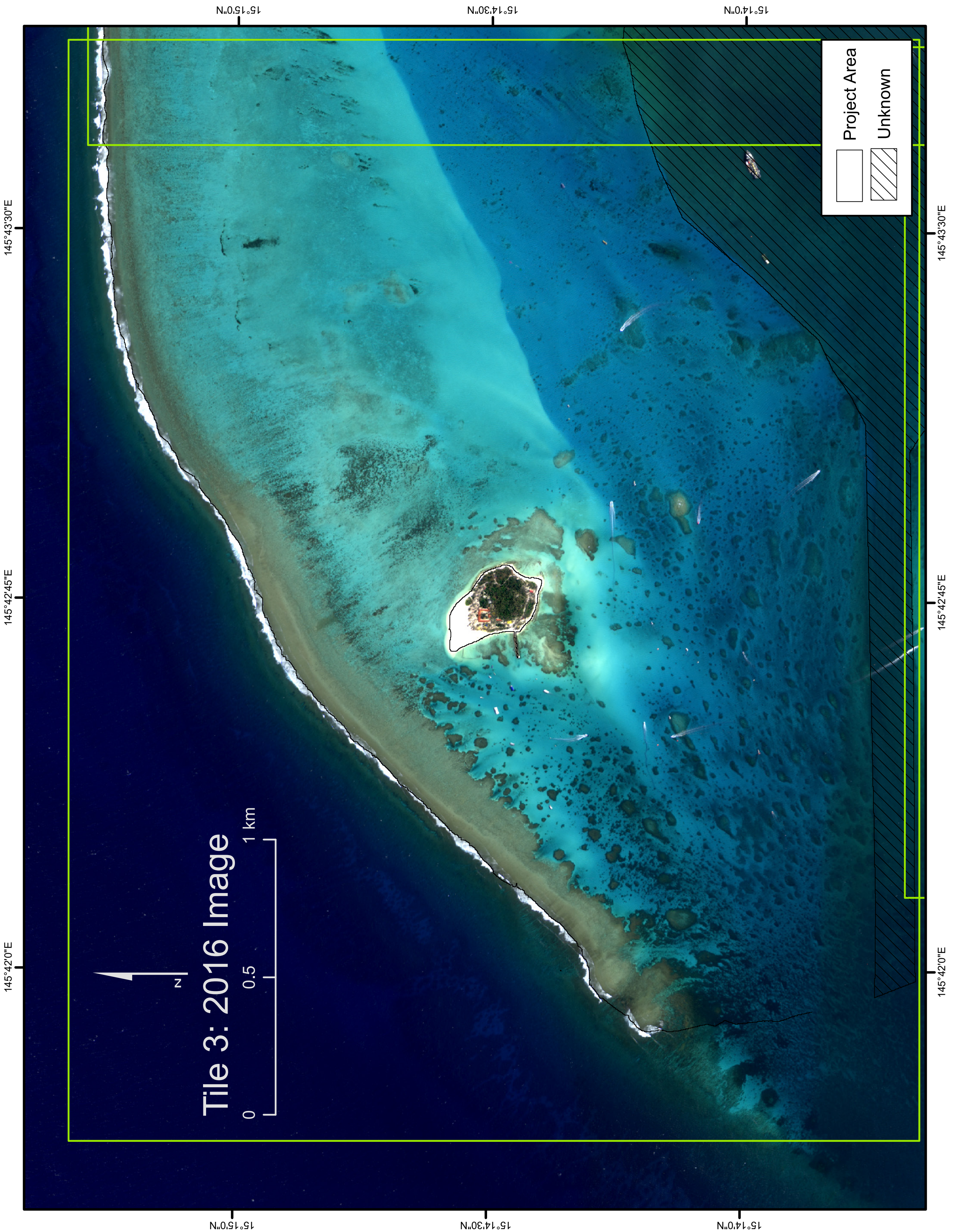


Tile 2: 2016 Image
Project Area
Unknown
0
0.5
1 km
N
15°15'0"N
15°14'30"N
15°14'0"N
145°45'45"E
145°45'0"E
145°44'15"E

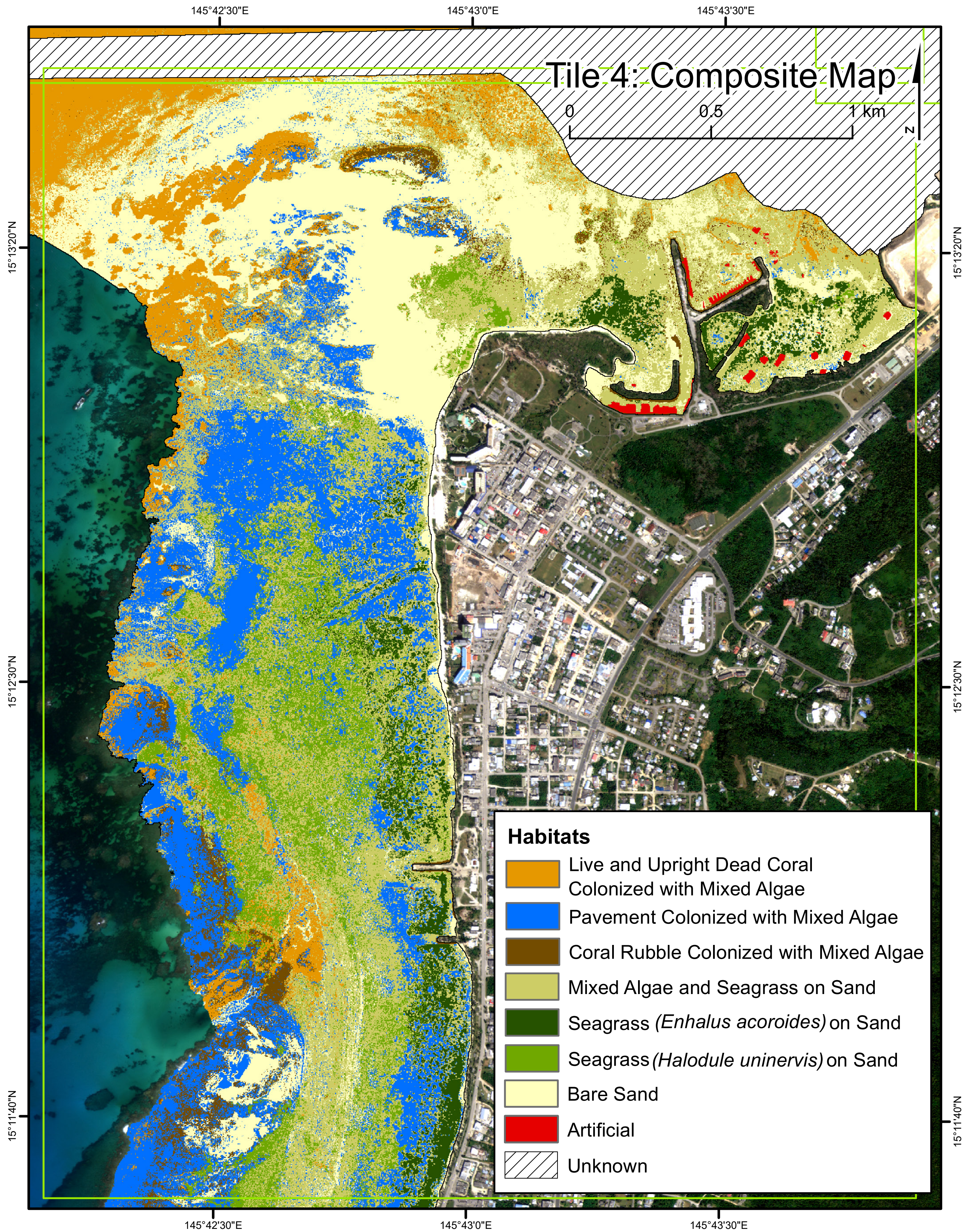
Appendix



Tile 3: Composite Map
0
0.5
1 km
N
Habitats
Live and Upright Dead Coral Colonized with Mixed Algae
Pavement Colonized with Mixed Algae
Coral Rubble Colonized with Mixed Algae
Mixed Algae and Seagrass on Sand
Seagrass(*Halodule uninervis*)on Sand
Bare Sand
Artificial
Unknown
15°15'0"N
15°14'30"N
15°14'0"N
145°43'30"E
145°42'45"E
145°42'0"E


Tile 3: 2016 Image
0
0.5
1 km
N
Project Area
Unknown
15°15'0"N
15°14'30"N
15°14'0"N
145°43'30"E
145°42'45"E
145°42'0"E

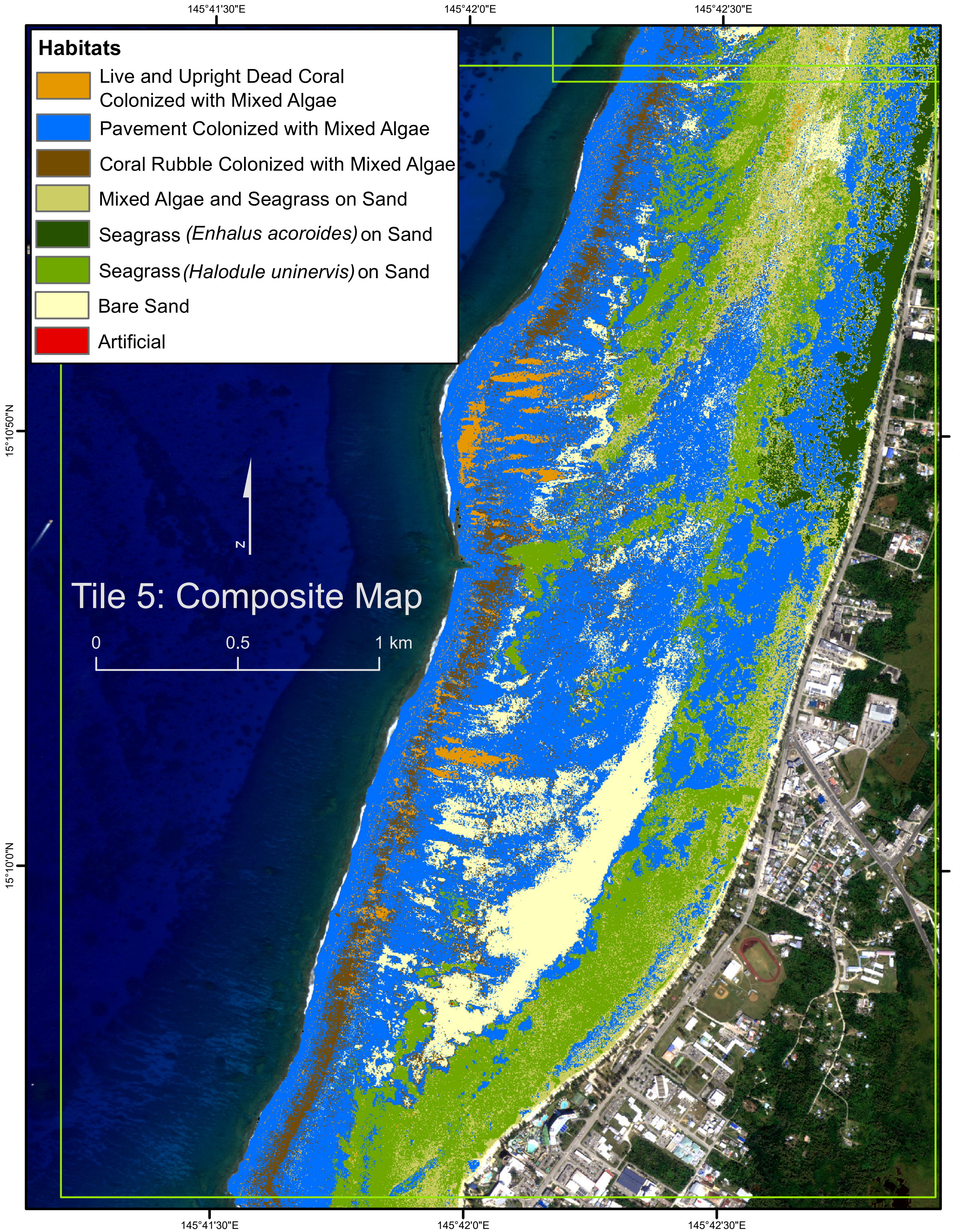
Appendix



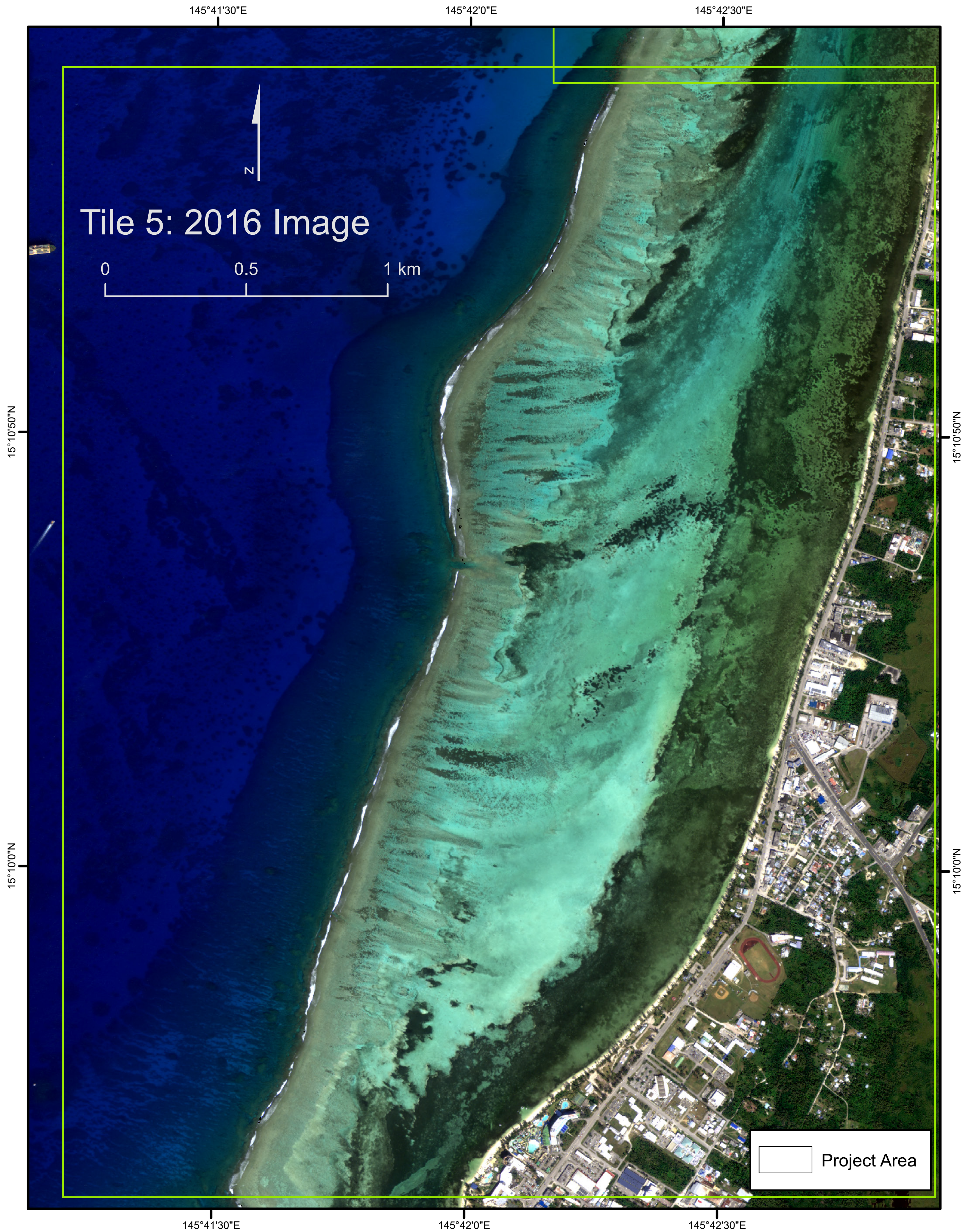
145°42'30"E
145°43'0"E
145°43'30"E
15°13'20"N
15°12'30"N
15°11'40"N
Tile 4: Composite Map
0
0.5
1 km
N
Habitats
Live and Upright Dead Coral Colonized with Mixed Algae
Pavement Colonized with Mixed Algae
Coral Rubble Colonized with Mixed Algae
Mixed Algae and Seagrass on Sand
Seagrass (Enhalus acoroides) on Sand
Seagrass (Halodule uninervis) on Sand
Bare Sand
Artificial
Unknown


145°42'30"E
145°43'0"E
145°43'30"E
Tile 4: 2016 Image
0
0.5
1 km
N
15°13'20"N
15°12'30"N
15°11'40"N
Project Area
Unknown

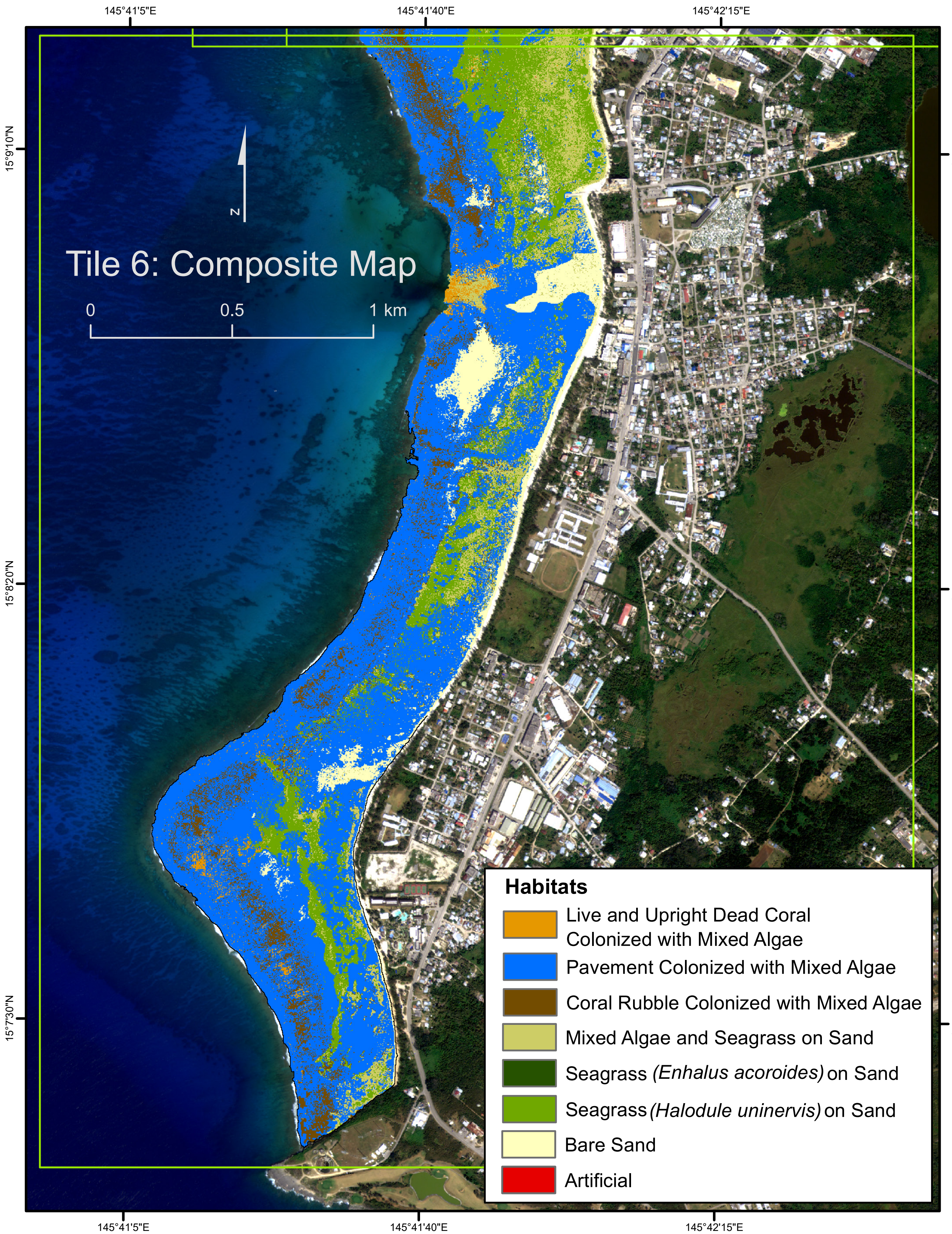
Appendix



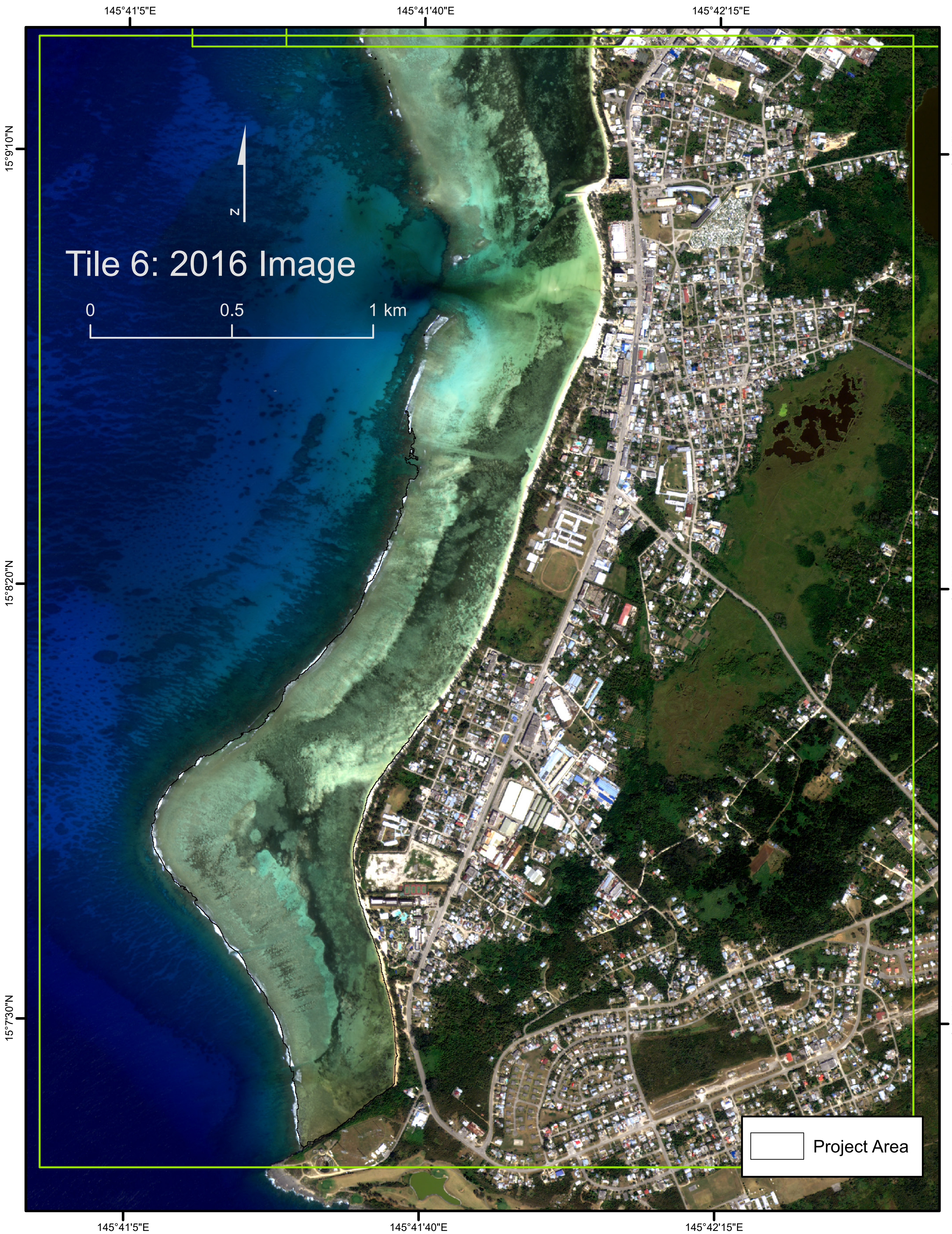
145°41'30"E
145°42'0"E
145°42'30"E
Habitats
Live and Upright Dead Coral Colonized with Mixed Algae
Pavement Colonized with Mixed Algae
Coral Rubble Colonized with Mixed Algae
Mixed Algae and Seagrass on Sand
Seagrass (Enhalus acoroides) on Sand
Seagrass (Halodule uninervis) on Sand
Bare Sand
Artificial
15°10'50"N
15°10'0"N
N
Tile 5: Composite Map
0
0.5
1 km


145°41'30"E
145°42'0"E
145°42'30"E
Tile 5: 2016 Image
0
0.5
1 km
N
15°10'50"N
15°10'0"N
Project Area

Appendix



145°41'5"E
145°41'40"E
145°42'15"E
15°9'10"N
15°8'20"N
15°7'30"N
N
Tile 6: Composite Map
0
0.5
1 km
Habitats
Live and Upright Dead Coral Colonized with Mixed Algae
Pavement Colonized with Mixed Algae
Coral Rubble Colonized with Mixed Algae
Mixed Algae and Seagrass on Sand
Seagrass (Enhalus acoroides) on Sand
Seagrass (Halodule uninervis) on Sand
Bare Sand
Artificial


145°41'5"E
145°41'40"E
145°42'15"E
15°9'10"N
15°8'20"N
15°7'30"N
N
Tile 6: 2016 Image
0
0.5
1 km
Project Area

U.S. Department of Commerce

Wilbur L. Ross, Jr., *Secretary*

National Oceanic and Atmospheric Administration

Benjamin Friedman, *Acting Under Secretary for Oceans and Atmosphere*

National Ocean Service

Russell Callender, *Assistant Administrator for National Ocean Service*