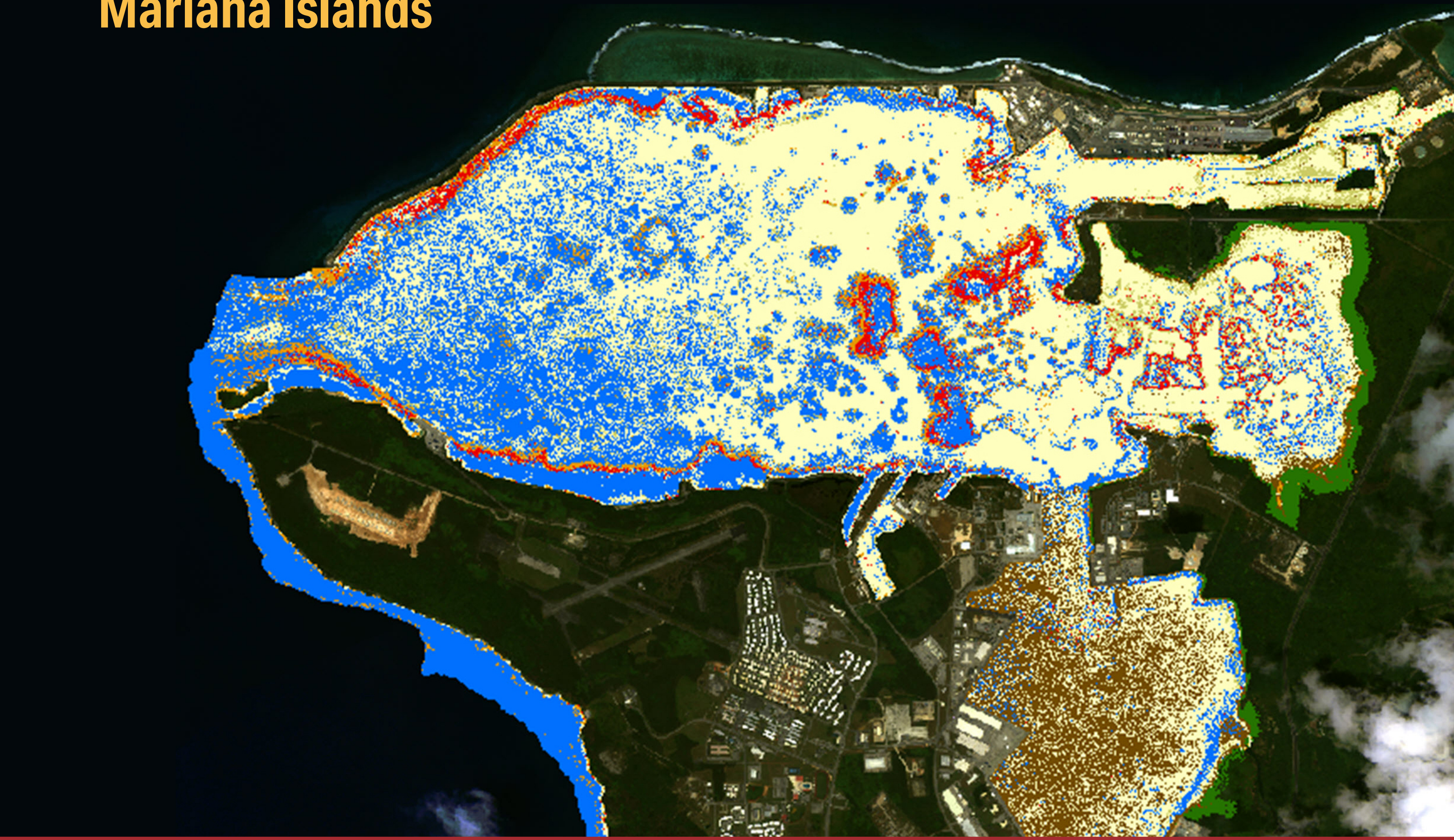
https://doi.org/10.25923/zwwa-h562

Characterizing Submerged Lands Around Naval Base Guam, Mariana Islands



Authors

Bryan Costa
Edward Sweeney

April 2024

NOAA TECHNICAL MEMORANDUM NOS NCCOS 329

NOAA NOS National Centers for Coastal Ocean Science

Suggested Citation

Costa, B., and Sweeney, E. (2024). Characterizing submerged lands around Naval Base Guam, Mariana Islands. NOAA Technical Memorandum NOS NCCOS 329. https://doi.org/10.25923/zwwa-h562

Acknowledgments

This work was conducted by NOAA National Centers for Coastal Ocean Science (NCCOS) in partnership with Naval Facilities Engineering Command Marianas (NAVFAC Marianas) and in consultation with the Guam territorial government (Division of Aquatic and Wildlife Resources, Bureau of Statistics and Plans), University of Guam (UoG), NOAA National Marine Fisheries Service (NMFS), and NOAA Office for Coastal Management (OCM). This work was funded by NAVFAC Marianas and with in-kind contributions from NOAA NCCOS. Local experts were consulted before, during, and after the mapping process in order to receive feedback on product accuracy and suggestions for customizing the products to meet their needs. Briefings and reviews included Andres Reyes, Jude Martinez, and Anjelica Perez (NAVFAC Marianas); Ashton Williams (Guam BSP); Peter Houk (UoG); David Burdick (UoG); Brent Tibbats (Guam DAWR); Laurie Raymundo (UoG); and Tom Schils (UoG). We collected and interpreted field data in collaboration with staff from the NOAA OCM (Marie Auyong). Captains Todd Genereux, Barnaby Acfalle, and BJ Acfalle shared their local knowledge and safely navigated us among field sites in Apra Harbor and Haputo ERA. Jennifer Kraus (NOAA NCCOS) processed environmental predictors. Chris Clement (NOAA NCCOS) color corrected the underwater photographs, and Jordan Pierce (CSS Inc., NOAA NCCOS) processed them into photomosaics. Ken Buja (NOAA NCCOS) created the online web map viewer. We also thank Dr. Matt Kendall (NOAA NCCOS), Dr. Arliss Winship (CSS Inc., NOAA NCCOS), and Jude Martinez (NAVFAC Marianas) for their constructive review comments on a draft of this document. The covers, layout, and formatting were completed by Sarah Hile (CSS Inc., NOAA NCCOS). Editorial content was improved by Ginger Shoup (CSS Inc., NOAA NCCOS). CSS employees were supported under NOAA Call Order GSA7-115-MSE.

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/project/characterizing-submerged-lands-around-navy-base-guam-cnmi/
https://coastalscience.noaa.gov/project/characterizing-benthic-habitats-in-haputo-ecological-reserve-area-guam/

Or direct questions and comments to:
NOAA/NOS/National Centers for Coastal Ocean Science
University of California Santa Barbara
Ocean Science Education Building 514, MC 6155
Santa Barbara, CA 93106-6155
bryan.costa@noaa.gov

Photography Credits

All images used in this report were taken by NOAA NCCOS unless otherwise indicated.

Disclaimer

Characterizing Submerged Lands Around Naval Base Guam, Mariana Islands

Prepared by

NOAA National Ocean Service
National Centers for Coastal Ocean Science
Marine Spatial Ecology Division
Silver Spring, MD

Authors

Bryan Costa[1]
Edward Sweeney[2]

[1] NOAA, National Ocean Service, National Centers for Coastal Ocean Science
[2] CSS, Inc. under contract to NOAA, National Ocean Service, National Centers for Coastal Ocean Science

April 2024

NOAA NOS NCCOS TECHNICAL MEMORANDUM 329

United States Department of Commerce

Gina M. Raimondo
Secretary

National Oceanic and Atmospheric Administration

Richard W. Spinrad, Ph.D.
Under Secretary of Commerce for Oceans and Atmosphere and NOAA Administrator

National Ocean Service

Nicole LeBoeuf
Assistant Administrator

Table of Contents

List of Figures and Tables

Executive Summary

Acropora coral, Guam. Credit: NOAA NCCOS

Guam (Guåhan) is home to tens of thousands of U.S. military personnel stationed at Naval Base Guam (NBG), Andersen Air Force Base (AAFB), and other installations. Over the last two decades, this military buildup and increased military activities have brought economic stimulus to Guam but also directly and indirectly displaced and impacted marine ecosystems in the area. These cumulative impacts were described in the Navy's environmental impact statement. Their integrated resource plan recommends potential ways to mitigate the impact of naval activities on Guam's ecosystems, including its coral reefs. To implement these strategies, the Naval Facilities Engineering Command Marianas (NAVFAC Marianas) requested new maps for submerged lands around NBG. No benthic habitat maps had been produced around Guam since 2010.

To meet this need, NOAA's National Centers for Coastal Ocean Science (NCCOS) collaborated with NAVFAC Marianas to develop detailed maps of the distribution of seafloor habitats, beginning with Apra Harbor and Haputo Ecological Reserve Area (ERA). Two new types of map products were made for these locations. The first product type describes the spatial distribution of seven substrate (e.g., sand) and 12 biological (e.g., seagrass) cover types. These classes were used to create 19 map layers, where each 2 × 2 m grid cell denotes the probability that a given substrate or cover type is present (0% to 100%). The second product was a classified map depicting the seven most common combinations of substrate and cover types in a single layer.

Both product types were created using machine learning models called boosted regression trees (BRTs) and boosted classification trees (BCTs). These approaches model complex, nonlinear relationships between a response (the presence or absence of 19 habitat types in underwater photographs at 236 sites) and predictors (43 spatial layers describing the marine environment). Performance and accuracy of these map products were evaluated using an independent set of underwater photographs from 241 validation sites. Results indicate that substrate and cover models and predictions were robust, since they had little bias, had a high probability of correctly predicting presences versus absences, and explained almost a quarter of the variation in the data. The classified habitat map was also high quality with an overall accuracy of 86.6%.

Over 21 km^2 of seafloor was characterized around NBG and Haputo ERA from 0- to 50-m depths. In Haputo ERA, "Pavement, Mixed Algae" was the most abundant habitat type, comprising 54.5% (1.1 km^2) of the area. The largest continuous patches were located on the fore reef along the coastline. Live coral was distributed throughout the ERA, with encrusting corals being most prevalent. As expected, no mangroves or seagrasses were present. Around NBG, "Sand, Bare" was the most abundant habitat type, comprising 42.3% (8.2 km^2) of the area. The largest continuous patches were in the eastern portion of Outer Apra Harbor, including Sasa Bay and south of Cabras Island. Live coral was common and most prevalent from San Luis Point around Point Udall to Acapa Point. *Halodule uninervis* seagrass was only documented outside Apra Harbor at two sites approximately 500 m north of Acapa Point. Mangroves were found only in nearshore areas in Sasa Bay and Inner Apra Harbor. No Endangered Species Act (ESA)–protected corals or nuisance algae (angel hair algae, *Chaetomorpha vieillardii*) were photographed in either project areas. One crown-of-thorns sea star (*Acanthaster planci*) was photographed in Haputo ERA. The prevalence of coral bleaching, coral paling, crown-of-thorns scarring, and marine debris was also very low (<1%, 0%, and <4%, respectively).

There are a wide range of applications for these new habitat predictions and maps, source imagery, and underwater photographs. In particular, these map products will be used by NAVFAC Marianas to inform their monitoring and management decisions and guide how best to minimize impacts to important habitats around NBG and Haputo ERA. In addition to supporting NAVFAC Marianas, these products may inform other local spatial-management decisions, 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.

An atlas showing the satellite imagery and classified habitat maps are provided at the end of this report. However, the best way to view and interact with these map products is by using geographic information system (GIS) software (e.g., ESRI ArcPro). These GIS products are archived at NOAA's National Centers for Environmental Information and are available through NCCOS's website at: https://coastalscience.noaa.gov/project/characterizing-submerged-lands-around-navy-base-guam-cnmi/ and https://coastalscience.noaa.gov/project/characterizing-benthic-habitats-in-haputo-ecological-reserve-area-guam/. For users without GIS software, an online map is also available on NOAA's GeoPlatform to view and interact with the habitat maps, source imagery, and field data: https://experience.arcgis.com/experience/7b6c0e7164234182985a89d5b5703475.

Live coral in project area, including Porites rus. *Credit: NOAA NCCOS*

Chapter 1 Introduction

Spanish Steps, Guam. Credit: NOAA NCCOS

1.1 Guam (Guåhan)

Guam is a U.S. island territory located approximately 1,700 miles south of Japan, and 3,500 miles west of Hawaii in the Mariana Archipelago. Given its proximity to the Coral Triangle, Guam has one of the most species-rich marine ecosystems among U.S. jurisdictions (Veron, 2000) with over 5,100 marine species, including 300 species of hard coral (Paulay, 2003; Burdick et al., 2008). This diverse coral reef ecosystem is vitally important to the fisheries (Allen and Bartram, 2008) and tourism economy of Guam, estimated at approximately $127 million per year (van Beukering et al., 2007). These ecosystems are also important to the history and culture of the Chamorro and Carolinian people and provide numerous non-economic goods and services to the residents of Guam (Allen and Bartram, 2008). Like many other populated islands in the Pacific, Guam's coral reef ecosystems are stressed by several threats, including land-based sources of pollution, overfishing, invasive species, marine heatwaves, ocean acidification (Burdick et al., 2008), and military activities on the island (U.S. Navy, 2022).

To mitigate these stresses, multiple marine reserves have been established to protect these resources. Three of these reserves are located within the project areas described here (Figure 1.1), including Sasa Bay Ecological Reserve Area (ERA), Orote Point ERA, and Haputo ERA. Haputo ERA is a 2-km^2 area on the west or leeward side of the island, west of Marine Corps Base Camp Blaz and Andersen Air Force Base (AAFB). This reserve is about 8 mi north of the capital city of Hagåtña, and hosts a patchwork of coral reefs and sandy beaches adjacent to karstic cliffs. Previous research has documented 944 species of marine animals in Haputo ERA, including 154 species of corals and 207 species of fishes (Amesbury et al., 2001; Donaldson et al., 2008; Burdick et al., 2008).

South of Haputo ERA is Orote Point ERA. This 0.5-km^2 ERA is located outside Apra Harbor along the southern shoreline of Orote Peninsula and was established to mitigate the construction of a naval wharf in Apra Harbor (Donaldson et al., 2008; U.S. Navy, 2012). The submerged lands in the Orote Point ERA are primarily carbonate pavement colonized by crustose coralline and turf algae. Approximately 1,252 species of marine animals have been recorded within the ERA, including 156 coral species (Paulay et al., 2001; Donaldson et al., 2008).

East of the Orote Peninsula is Sasa Bay ERA, which is located inside Outer Apra Harbor. Apra Harbor generally and the Sasa Bay reserve in particular are well protected from ocean swells because of the Orote Peninsula to the south and Glass Breakwater to the north. The only opening to the ocean is a 1-km) wide channel to the west, providing unique conditions for coral, sponge (*Ianthella basta*), and mangrove habitats, which are found nowhere else on the island. This protection is also the reason Apra Harbor is a major port, with multiple marinas, industries, and U.S. military wharfs lining its shore. The Navy operates three wharfs in Outer Apra Harbor, and several other wharfs in Inner Apra Harbor. The largest of these installations is U.S. Naval Base Guam (NBG), located in Inner Apra Harbor.

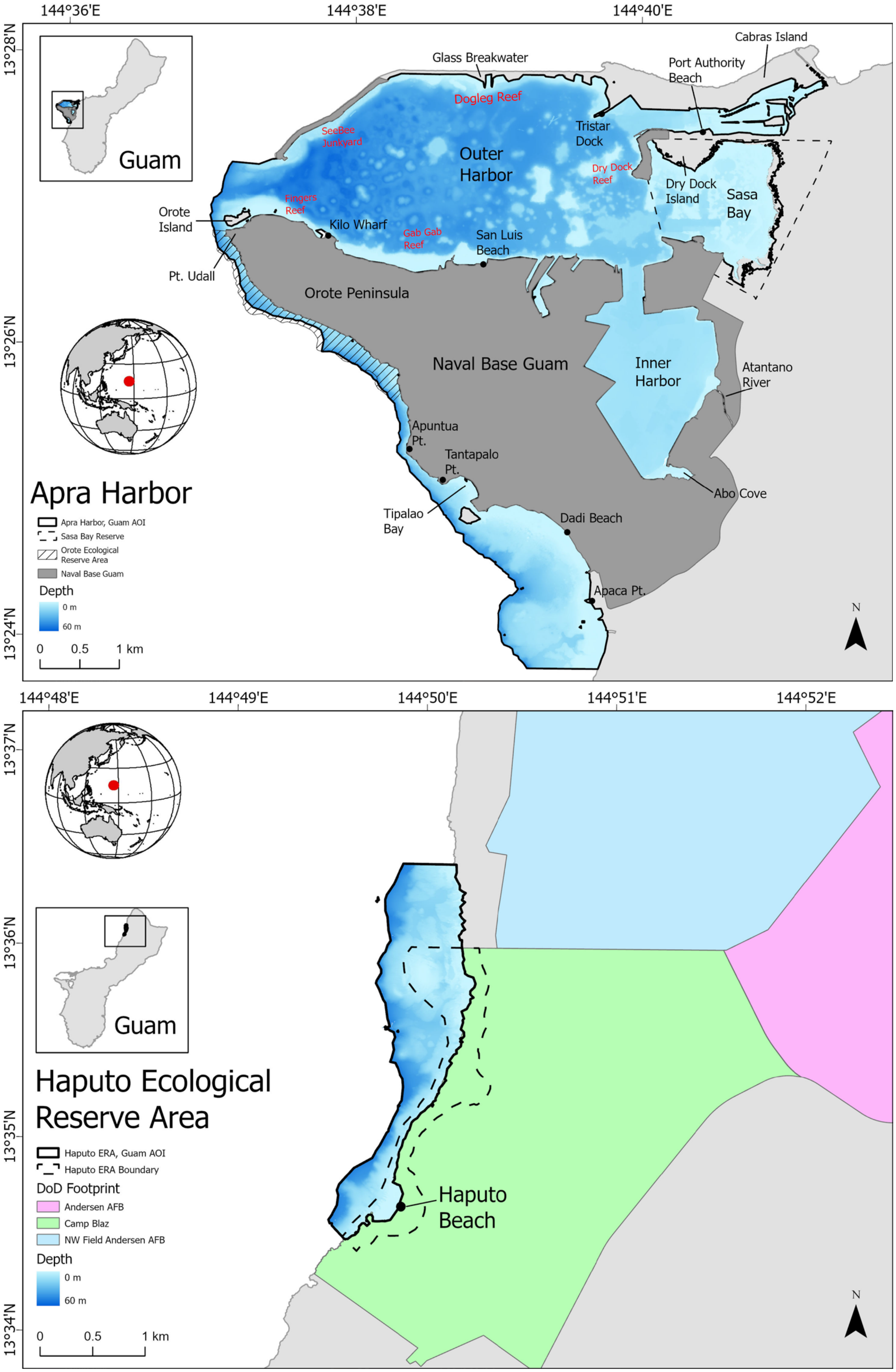


Figure 1.1. Key geographic features and place names around Apra Harbor (top) and Haputo Ecological Reserve Area (ERA) (bottom). AFB = air force base; AOI = area of interest.

1.2 U.S. Military Presence on Guam

Guam is home to tens of thousands of U.S. military personnel stationed at NBG, AAFB, and other installations on the island (Figure 1.2). The majority of these personnel arrived after 2008, although Apra Harbor has been a strategic location for the U.S. military since before World War II (U.S. Navy, 2012, 2022). Over the last two decades, population growth and the associated dredging and construction of new facilities has brought economic stimulus to Guam but also directly and indirectly displaced and impacted marine species and coral reefs in the area (Marler and Moore, 2011). These cumulative impacts were described in the Navy's environmental impact statement in 2010 (U.S. Navy, 2010). This environmental impact statement also outlined proposed mitigation strategies and preferred alternatives for the region.

Since 2008, Apra Harbor has and is being used by the Navy for a variety of training exercises and activities (U.S. Navy, 2012, 2022), which have the potential to impact nearby marine organisms and ecosystems. Like other federal agencies, the Navy is responsible for compliance with all territorial and federal environmental and natural resource laws and regulations that apply to the marine environment. This list includes (but is not limited to) the National Environmental Policy Act, the Marine Mammal Protection Act, the Endangered Species Act (ESA), the Magnuson-Stevens Fishery Conservation and Management Act/ Sustainable Fisheries Act, the Sikes Act (10 U.S.C. 670), and Executive Order 13089 on Coral Reef Protection. Naval Facilities Engineering Command Marianas (NAVFAC Marianas) ensures compliance and manages the natural resources for Marine Corps Base Camp Blaz, AAFB, NBG, and all submerged lands adjacent to its holdings from the shoreline out 3 nautical miles (U.S. Navy, 2022). In total, NAVFAC Marianas manages approximately 345 km^2 of submerged lands around the island of Guam, which encompass Haputo, Orote, and Sasa Bay ERAs (U.S. Navy, 2022).

To guide compliance, the Navy developed a resource management plan—called an integrated natural resources management plan (INRMP)—for the Marianas Region in collaboration with various federal and territorial agencies, including National Oceanic and Atmospheric Administration (NOAA) (U.S. Navy, 2012, 2022). This plan is updated no less than every five years, and its goal is "to provide for the restoration and enhancement of habitats for native species including listed species over the long-term in a manner that is consistent with the military mission" (U.S. Navy, 2012, 2022). To meet this goal, the INRMP recommends potential ways for the Navy and NAVFAC Marianas to mitigate the impact of naval activities on Guam's ecosystems, including coral reefs. Mitigation strategies include: (1) establishing long-term ecosystem-based management plans to maintain submerged lands and ERAs, (2) implementing strategies to monitor health, reduce threats, and enhance coral reefs, (3) implementing management actions to protect and improve the status of marine species of the greatest concern, and (4) enhancing management through the use of geographic information system (GIS) information, development of cooperative partnerships, and education programs.

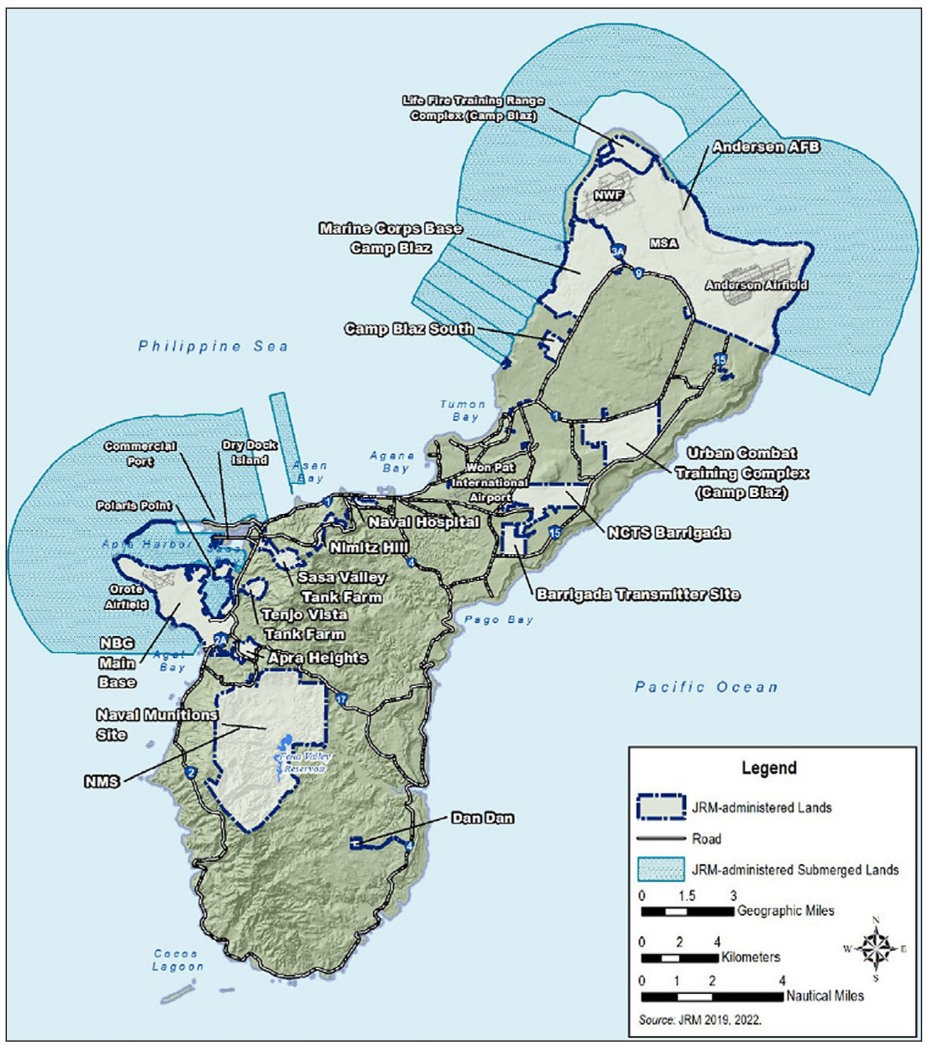


Figure 1.2. Location of Joint Region Marianas (JRM) sites on Guam. Figure adapted from U.S. Navy (2022). AFB = Air Force Base; NBG = Naval Base Guam; NMS = Naval Munitions Site; NCTS = Naval Computer and Telecommunications Station; NWF = Northwest Field; MSA = Munitions Storage Area.

1.3 Characterizing Submerged Lands Around NBG

To implement mitigation strategies in INRMP, NAVFAC Marianas requested new benthic habitat maps for submerged lands around Naval Base Guam, beginning with Apra Harbor (including Sasa Bay and Orote ERAs) and Haputo ERA. These maps will help NAVFAC Marianas evaluate different management scenarios, monitor changes, and validate the efficacy of mitigation strategies over time. A few benthic habitat maps have been produced around Guam since 2001. Habitats and biota were first mapped and characterized in 2001 (Amesbury et al., 2001; Paulay et al., 2001) as part of a baseline survey for Haputo and Orote ERAs. These surveys resulted in a coarse depiction of macrohabitats. In 2005, NOAA National Centers for Coastal Ocean Science (NCCOS) mapped all shallow-water (<30 m) habitats around Guam (NOAA NCCOS, 2005) as part of its comprehensive initiative to characterize U.S. coral reef ecosystems (NOAA, 2002). This habitat map used a standard minimum mapping unit of 4,047 m^2 (i.e., habitat features smaller than 4,047 m^2 were not digitized and classified). It also used a hierarchical classification scheme that included attributes for reef zone, geomorphological structure, and density of biological cover (Figure 1.3; NOAA NCCOS, 2005). Habitats were also digitized using the same imagery at a finer spatial scale by Burdick (2005). In 2010, the Pacific Islands Benthic Habitat Mapping Center developed a hard and soft seafloor substrate map classified for Apra Harbor (PIBHMC, 2010).

Since these last mapping products were completed, several natural disasters have potentially altered benthic habitats around the island. Perhaps most notable, four near-consecutive coral bleaching events (2013, 2014, 2016, and 2017) occurred around Guam, with up to 80% of corals bleaching at some locations (NOAA CRCP, 2018). These events led to live coral cover declining by 37% at sites along the leeward coast and by 34% at shallow seaward slope sites around the island (Raymundo et al., 2019). These events severely impacted the condition of Guam's leeward reefs, including in Apra Harbor and Haputo ERA, causing their substantial decline (NOAA CRCP, 2018). These coral bleaching events, coupled with the ongoing impacts from the military buildup on Guam, have increased the need for an updated map of benthic habitats around Guam.

Here, the goal was to produce new, highly detailed maps of benthic habitats around NBG and Haputo ERA for NAVFAC Marianas in 0- to 50-m depths (Figure 1.1). Recent advancements in remote sensing, machine learning models, and cloud-based computational power has enabled a new generation of habitat map products. The result is a dramatic increase in map detail, from approximately 4,047-m^2 polygons that were hand digitized to a 4-m^2 grid that was attributed using machine learning models. Here, machine learning models called boosted regression trees (BRTs) and boosted classification trees (BCTs) were applied to model complex, nonlinear relationships between a response (the presence or absence of 19 habitat types and seven habitat classes in underwater photographs) and predictors (43 spatial layers describing the marine environment). The resulting products included 19 individual substrate and biological cover spatial predictions and a single classified benthic habitat map for the project area. The thematic and spatial accuracy of these products was qualitatively evaluated by local experts and quantitatively measured using independent field data. This report describes the methods used and accuracy of these products, the broad spatial distribution of habitats in the project area, and the potential applications of these maps for marine science and management decisions in Guam.

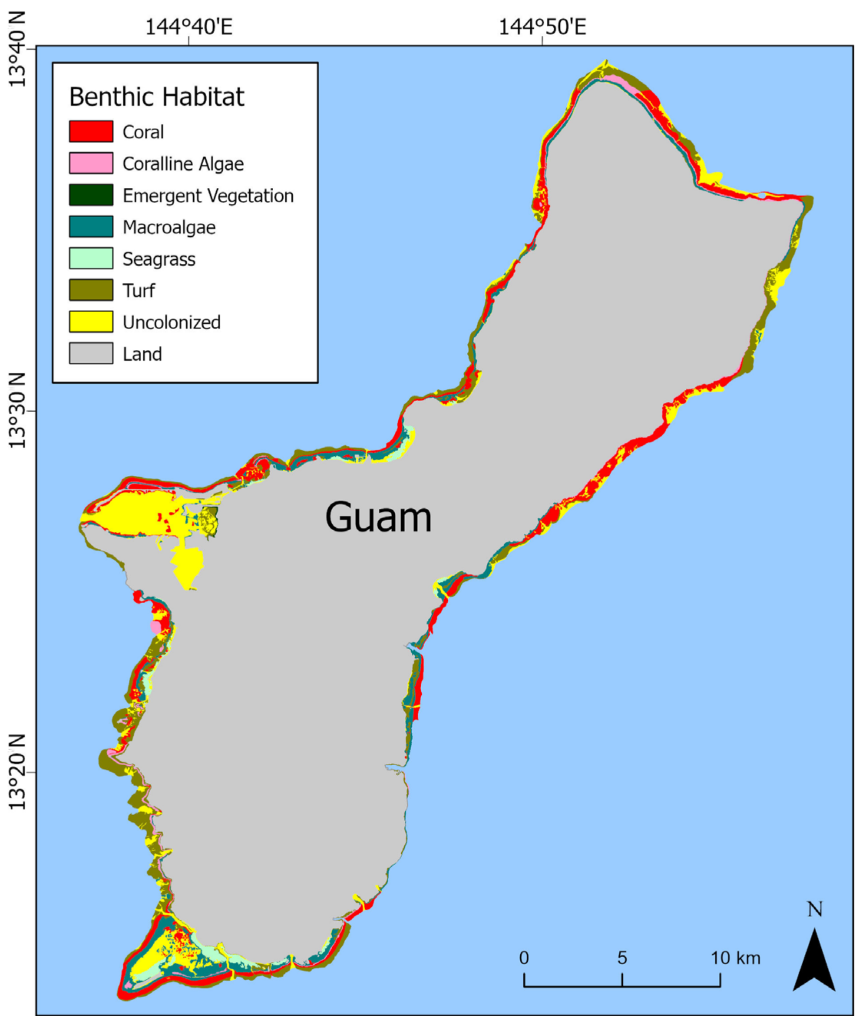


Figure 1.3. Benthic features mapped by NOAA NCCOS (2005) around Guam.

Chapter 2 Methods

Live coral in Apra Harbor, including Porites rus. *Credit: NOAA NCCOS*

Several steps were needed to map and characterize habitats inside Haputo ERA, Apra Harbor, and along the coast between the Apra Harbor mouth and Dadi Beach (Figure 1.1). This section describes the steps used during map development, including: (1) customizing a habitat classification scheme; (2) processing environmental variables including satellite, topographic, acoustic, and geographic predictors; (3) collecting and annotating underwater photographs; (4) creating habitat predictions and a classified map using two spatial predictive modeling techniques—BRTs and BCTs, and (5) assessing the performance and accuracy of the habitat predictions and the classified map.

2.1 Benthic Habitat Classification Scheme

A habitat classification scheme allows scientists to systematically group benthic features based on their ecological characteristics. The classification scheme used here was developed by reviewing the previous habitat classifications applied in Guam and in the Northern Mariana Islands (Cloud, 1959; NOAA NCCOS, 2005; Houk and van Woesik, 2008; Kendall et al., 2017) and consulting with local scientists and managers about their informational needs. The scheme is based on 19 benthic habitats including seven substrate (e.g., sand) and 12 biological (e.g., turf algae) cover types found around Guam (Table 2.1). These 19 habitats guided collection of field data and development of two types of map products. The first products were spatial predictions for 19 substrate and biological cover types inside the project area. This modeling resulted in 19 individual map layers with each 2 × 2 m grid cell denoting the mean probability (from 0% to 100%; averaged from 100 bootstrapped model iterations) that a given habitat type is present.

The second product was a classified habitat map depicting commonly occurring combinations of substrates and cover types. It was developed using the 19 probability of occurrence predictions described above. Habitats in the classified map were defined based on cluster analysis of the field data (R Core Team, 2022; Maechler, 2023; performed in R using the agnes, diana, and hclust functions in the cluster package). Clustering is a statistical technique used to identify groups of similar objects based on multiple attributes. Six clustering techniques were tested (average, single, complete, Ward, divisive, and McQuitty) using percent cover for each substrate and cover type as the input. Percent cover was grouped into five, six, seven, and eight clusters at all training and validation sites. The technique and number of clusters with the highest agglomerative coefficient was selected (i.e., Ward, seven clusters, agglomerative coefficient = 0.99). The resulting seven habitat classes (Table 2.2) were the basis of the new benthic habitat map. These classes were also translated into the Coastal and Marine Ecological Classification Standard (CMECS) (Tables 2.1 and 2.2; CMECS, 2023). NOAA is required to use CMECS by the Federal Geographic Data Committee.

Table 2.1. Substrate and cover types used to develop individual benthic habitat predictions. Equivalent Coastal and Marine Ecological Classification Standard (CMECS) classifications are suggested.

		Habitats	Definition	CMECS IDs	CMECS Class
Substrate	1	Live Coral Reef (All Species)	Presence of live coral reef. Comprising live, upright hermatypic (reef-building) hard corals, including all hard coral species	g2.5, 2.2.1, 2.1.2	Geoform – Shallow/Mesophotic Coral Reef, Substrate – Coral Reef Substrate, Biotic – Shallow/Mesophotic Coral Reef Biota
Substrate	2	Upright Dead Coral Reef	Presence of dead hard coral reef that is still upright	g2.5, 2.2.1	Geoform – Shallow/Mesophotic Coral Reef, Substrate – Coral Reef Substrate
Substrate	3	Rock	Presence of non-biogenic rock	1.1	Substrate – Rock Substrate
Substrate	4	Pavement	Presence of flat, low-relief or sloping solid carbonate rock with little or no fine-scale rugosity that is covered with algae, coral, zoanthids or other sessile vertebrates that are dense enough to partially obscure the underlying surface.	g1.44, 2.2.1	Geoform – Pavement Area, Substrate – Coral Reef Substrate
Substrate	5	Rubble	Presence of dead, unstable coral rubble often colonized with filamentous or other macroalgae. This habitat often occurs landward of well-developed reef formations.	2.2.2	Substrate – Coral Rubble
Substrate	6	Sand	Presence of coarser sediment (with grain sizes between 0.0625 and 2 mm) typically found in areas exposed to currents or wave energy.	1.2.2.2	Substrate – Sand
Substrate	7	Mud	Presence of fine sediment (with grain sizes less than 0.0625 mm) typically found in areas with little wave energy.	1.2.2.5	Substrate – Mud
Biological Cover	1	Mangrove	Presence of various species of mangrove	2.8.1.4	Biotic – Tidal Mangrove Forest
Biological Cover	2	Live Coral (Branching)	Presence of live branching coral	2.1.2.1	Biotic – Branching Coral Reef
Biological Cover	3	Live Coral (Encrusting)	Presence of live encrusting coral	2.1.2.3	Biotic – Encrusting Coral Reef
Biological Cover	4	Live Coral (Foliose)	Presence of live foliose coral	2.1.2.4	Biotic – Foliose Coral Reef
Biological Cover	5	Live Coral (*Porites rus*)	Presence of *Porites rus* coral	2.1.2.1.4	Biotic – Branching Porites Reef
Biological Cover	6	Seagrass (*Halodule uninervis*)	Presence of *Halodule uninervis* seagrass	2.5.2.1	Biotic – Seagrass Bed
Biological Cover	7	Algae (Crustose Coralline)	Presence of crustose coralline algae (CCA)	2.5.1.3	Biotic – Coralline/Crustose Algal Bed
Biological Cover	8	Algae (*Halimeda* spp.)	Presence of *Halimeda* algae	2.5.1.1.2	Biotic – *Halimeda* Communities
Biological Cover	9	Algae (Turf)	Presence of turf algae	2.5.1.8	Biotic – Turf Algal Bed
Biological Cover	10	Algae (Other)	Presence of mixed macroalgae not already identified	2.5.1	Biotic – Benthic Macroalgae
Biological Cover	11	Sponges	Presence of sponges, primarily *Ianthella basta*	2.2.1.21	Biotic – Attached Sponges
Biological Cover	14	Bare	No biological cover	No Equivalent	NULL

Mangroves in Sasa Bay, Guam. Credit: David Burdick (NOAA)

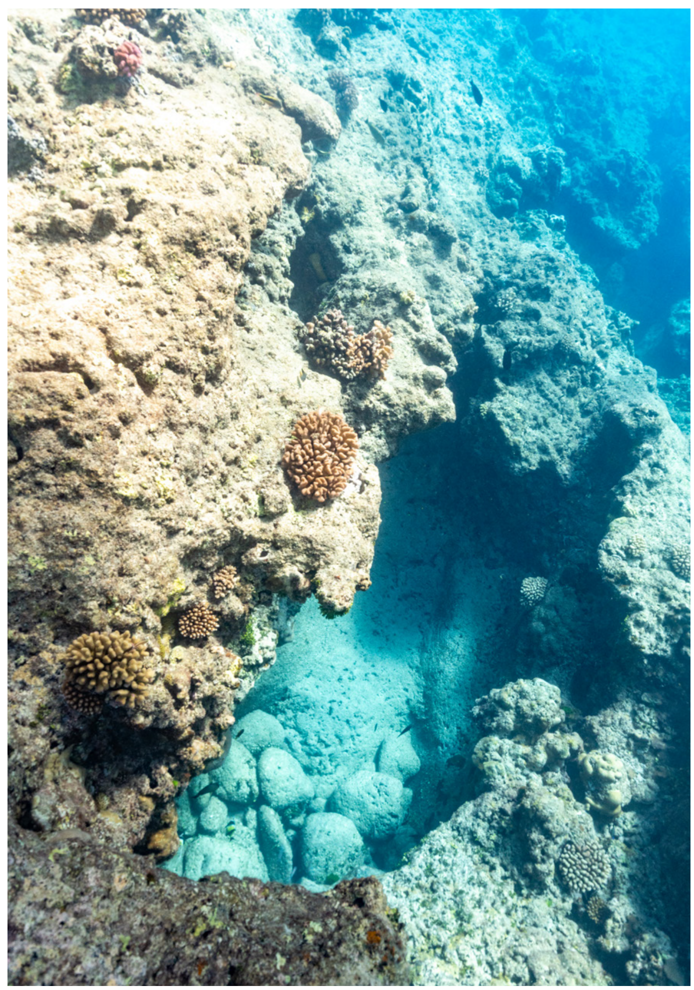
Table 2.2. Commonly co-occurring substrate and cover types used to create the classified benthic habitat map. Equivalent Coastal and Marine Ecological Classification Standard (CMECS) classifications are suggested.

Code	Habitat	Definition	CMECS ID (Geoform or Substrate)	CMECS Class (Geoform or Substrate)	CMECS ID (Biotic)	CMECS Class (Biotic)
1	Live Coral Reef, Live Coral	Majority live coral reef with live hard coral. The remaining percent is primarily dead reef with macroalgae.	g2.5, 2.2.1	Geoform - Shallow/Mesophotic Coral Reef, Substrate - Coral Reef Substrate	2.1.2	Shallow/Mesophotic Coral Reef Biota
2	Pavement, Mixed Algae	Majority pavement primarily covered with turf algae. The remaining percent is mixed proportions of hard and soft substrates with macroalgae.	g1.44, 2.2.1	Geoform - Pavement Area, Substrate - Coral Reef Substrate	2.5.1	Benthic Macroalgae
3	Sand, Mixed Algae	Majority sand that is bare or covered with macroalgae. The remaining percent is mixed proportions of hard substrates with macroalgae.	1.2.2.2	Substrate - Sand	NULL, 2.5.1	No equivalent, Benthic Macroalgae
4	Upright Dead Coral Reef, Mixed Algae	Majority upright dead coral reef covered with primarily turf algae. The remaining percent is mixed proportions of hard and soft substrates with some live coral and macroalgae.	g2.5, 2.2.1	Geoform - Shallow/Mesophotic Coral Reef, Substrate - Coral Reef Substrate	2.5.1	Benthic Macroalgae
5	Sand, Bare	>90% Sand that is >90% bare	1.2.2.2	Substrate - Sand	NULL	No Equivalent
6	Mud, Bare	>90% Mud that is >90% bare	1.2.2.5	Substrate - Mud	NULL	No Equivalent
7	Mud, Mangrove	>90% Mud that is >90% mangroves	1.2.2.5	Substrate - Mud	2.8.1.4	Tidal Mangrove Forest

In addition to the habitats above, other biological organisms were identified and observations made from the underwater photographs at the request of local managers. These organisms and observations were not modeled because either: (1) their prevalence was too low (<3%) to develop model predictions or (2) their model predictions did not meet the minimum performance thresholds (i.e., receiver operating characteristic [ROC] area under the curve [AUC] ≥0.7 and percent deviance explained [PDE] >0). Organisms that were not modeled specifically included: cyanobacteria, angel hair algae (*Chaetomorpha vieillardii*), mushroom corals, soft corals, fire corals, crown-of-thorns sea stars (*Acanthaster planci*), and species listed under the ESA (*Acropora globiceps*, *Isopora palifera*, *Acropora retusa*, and *Seriatopora aculeata*). The presence of coral bleaching, coral paling, crown-of-thorns scarring, and marine debris were also observed and documented. The spatial distributions of these organisms and observations are reported in the results.

2.2 Predictor Data

Forty-three environmental variables were used to create the model predictions for individual substrate and cover types. These predictors were divided into four broad categories: 23 spectral variables derived from satellite imagery, three geographic variables based on relative location, two predictors from acoustic backscatter, and 15 topographic surfaces derived from the depth and elevation surface.



Benthic habitats in Haputo ERA. Credit: NOAA NCCOS

2.2.1 Spectral Predictors

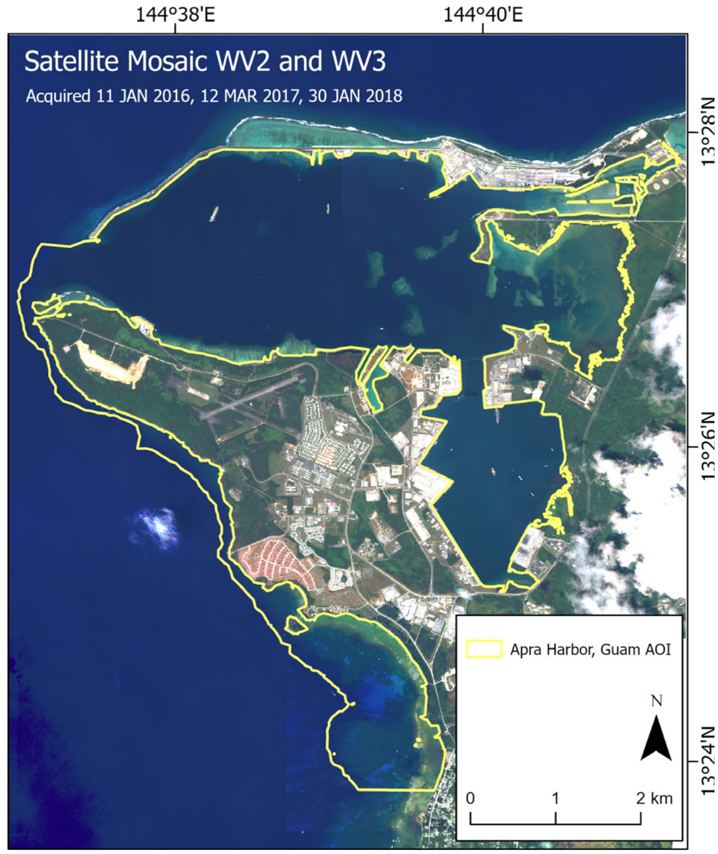
Twenty-three of the 43 environmental predictors were derived from WorldView-2 (WV2) and WorldView-3 (WV3) satellite images. The WV2 sensor collects eight multispectral bands in the visible near infrared at 2 × 2 m spatial resolutions, and the WV3 sensor collects eight multispectral bands in the visible near infrared at 1.4 × 1.4 m spatial resolutions (upsampled to 2 × 2 m). These images were acquired on 11 January 2016, 12 March 2017, and 30 January 2018 in Apra Harbor; and 18 February 2020 in Haputo ERA. The satellite scenes were very high quality but contained some artifacts due to the presence of ships, clouds, ship wakes, and turbidity. Additionally, no one scene covered the entire project area. Consequently, these scenes were mosaicked to remove artifacts and create a single satellite image for the project area (Figure 2.1).

To correct geometric distortions, spectral bands were orthorectified using 40 ground control points (GCPs) and a digital elevation model (USGS, 2010; performed in PCI OrthoEngine). The final orthorectified image (orthoimage) was georeferenced 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 16 GCPs collected using a Trimble GeoXH 6000 global positioning system (GPS) receiver from 8 May to 13 May 2022. GCPs were evenly distributed in the project areas and positioned on features that were clearly identifiable in the imagery, such as street intersections, parking lots, crosswalks, and other low-profile objects with distinct edges. Raw GPS data were post-processed and differentially corrected with Trimble Pathfinder Office software and data from the Mariana Island Continuously Operating Reference System station.

The combined root mean square error (RMSE) is 6.1 m for the Apra Harbor and Haputo satellite mosaic. This indicates that pixels in the mosaic were on average ± 6.1 m (three pixels) from their true location. This positional uncertainty was taken into account when evaluating the accuracy of the classified benthic habitat map. The orthoimages and mosaic were also corrected for changing atmospheric and water column conditions (Lyzenga, 1978; Mumby and Edwards, 2000; performed using ENVI 5.7: THOR atmospheric correction tool and R code). These processes resulted in 15 atmospheric- and water column–corrected band pairs (Figure A1).

2.2.2 Topographic Predictors

Seafloor depth and topography are known to be useful predictors for marine habitats, such as sand, pavement, and coral reefs. Elevation and topography are similarly useful predictors for estuarine habitats, such as mangroves.



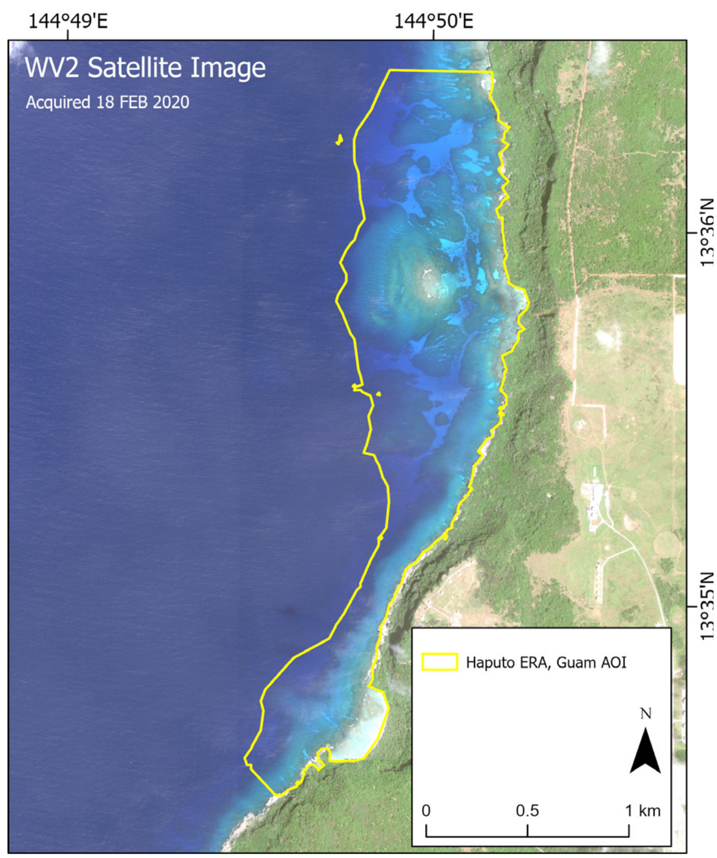


Figure 2.1. Maps depicting the WorldView-2 (WV2) and WorldView-3 (WV3) images acquired for Apra Harbor (top) and Haputo Ecological Reserve Area (ERA) (bottom) used to create the habitat predictions. AOI = area of interest.

Consequently, a single elevation and depth surface was created by merging data from a 2017 R2 Sonic multibeam echosounder survey (HDR and CSA Ocean Sciences, 2017) with a 2020 Leica Chioptera-4 lidar survey (NOAA NGS, 2020). Satellite-derived bathymetry was also developed from the above satellite mosaic (Kuhn et al., 2023; performed in R caret package) to fill in remaining data gaps as needed. The resulting elevation and bathymetry surface had a 2 × 2 m spatial resolution and was referenced to mean lower low water tidal datum.

From this single elevation and bathymetry surface, 14 topographic predictors were derived: (1) Elevation and Depth (Standard Deviation), (2) Aspect, (3) Aspect Northness (Cosine), (4) Aspect Eastness (Sine), (5) General Curvature, (6) Longitudinal Curvature, (7) Planform Curvature, (8) Profile Curvature (Evans), (9) Profile Curvature (Zevenbergen and Thorne), (10) Total Curvature, (11) Rugosity, (12) Slope, (13) Slope Rate of Change, and (14) Surface Area (Figure A2). Each topographic surface was calculated using the default 3 × 3 pixel neighborhood (Hijmans, 2023a; performed using R code, raster package). Multiple topographic metrics (e.g., curvature) were derived to explore which surfaces most uniquely described the complexity of the seafloor. Highly correlated surfaces (Spearman rank $\rho \geq 0.9$ or $\rho \leq -0.9$) were removed later in the modeling process.

2.2.3 Acoustic Predictors

While bathymetry is important for identifying benthic habitats, depth and topography alone do not capture the complexity, texture, and composition of seafloor substrates and habitats. Acoustic backscatter can help fill this data gap and provide additional critical information about the hardness and roughness of the seafloor. Given the utility of backscatter, two acoustic backscatter surfaces were also included in the modeling process as environmental predictors. These surfaces were developed from the 2017 multibeam echosounder data in Outer Apra Harbor (HDR and CSA Ocean Sciences, 2017) and from a 2001 sidescan survey in Inner Apra Harbor (NOAA, 2001; Figure A3). They only included areas deeper than 5 m in Outer and Inner Apra Harbor. The effect of this data gap on the modeling process is discussed in the results section.

2.2.4 Geographic Predictors

Three geographic predictors were used to account for spatial variation in benthic habitats that was not explained by the spectral, topographic, or acoustic predictors. These included latitude (y), longitude (x), and distance to shore (Figure A4; performed using the ArcGIS Pro 2.8 Euclidean Distance tool). The shoreline was extracted from NOAA's previous benthic habitat map (NOAA NCCOS, 2005).

2.3 Field Data

2.3.1 Field Data Acquisition

NOAA NCCOS collected underwater photographs at 477 sites (Figure 2.2) between 2 May and 12 May 2022 throughout the project areas. One portion of this dataset (n = 236 sites) was used to train and tune the habitat models by correlating the response (observed substrate and cover types) with the values of the predictors (environmental layers). Locations of these training sites were selected visually beforehand (using the predictors) and spread out across the project areas to include the full range of habitats, depths, and environmental conditions found in Apra Harbor and Haputo ERA. The rest of the dataset (n = 241 sites) was used to validate the performance of the habitat models and evaluate the accuracy of the classified habitat map. Locations of these validation sites were randomly stratified based on an existing map of geomorphological structure types around Guam (NOAA NCCOS, 2005). The total number of sites was determined by the amount of allocated funding and the availability of NCCOS staff to conduct the field work.

When in the field, the process for collecting overlapping, underwater photographs was identical at each site. A handheld Garmin 76 GPS unit was used to navigate to each site via a small boat or paddle board. At sites accessed by small boat, a drop camera system (Figure 2.3) was lowered to within 1.5 to 2 m of the seafloor. This drop camera system was designed by NCCOS, and it included: (1) a downward-looking Sony α7 IV mirrorless 24 MP full-frame digital single-lens reflex (DSLR) camera, (2) an oblique-looking GoPro HERO10 Black in a Spot X Squid housing, (3) two green lasers spaced 10 cm apart, and (4) a Blueprint Subsea SeaTrac ultra-short baseline (USBL) transponder.

The downward-looking Sony camera collected still photographs of the seafloor every 0.5 s. The camera settings included: focal lengths = 18 mm, white balance = auto, and shutter speed = 1/200 s. Both jpeg and .RAW files were recorded. Each photograph covered an estimated 4 m^2 of the seafloor (at 1.6 m altitudes) and overlapped other photographs by 60%–80%. The lasers provided a measurement scale (10 cm) and were visible in some but not all of the photographs. The Trimble GeoXH GPS provided the location of the vessel every 1 s, and the USBL transponder provided the location and depth (XYZ) of the drop camera every 5 s. The Trimble GPS antenna was positioned directly over the USBL transponder pole to minimize lever arm offsets. For sites shallower than 2 m, the Sony camera was attached to a paddle board with the Trimble GPS directly over the camera. Photographs and GPS data were acquired while moving the paddle board around the site, imaging a minimum of a 4-m^2 area on the seafloor.

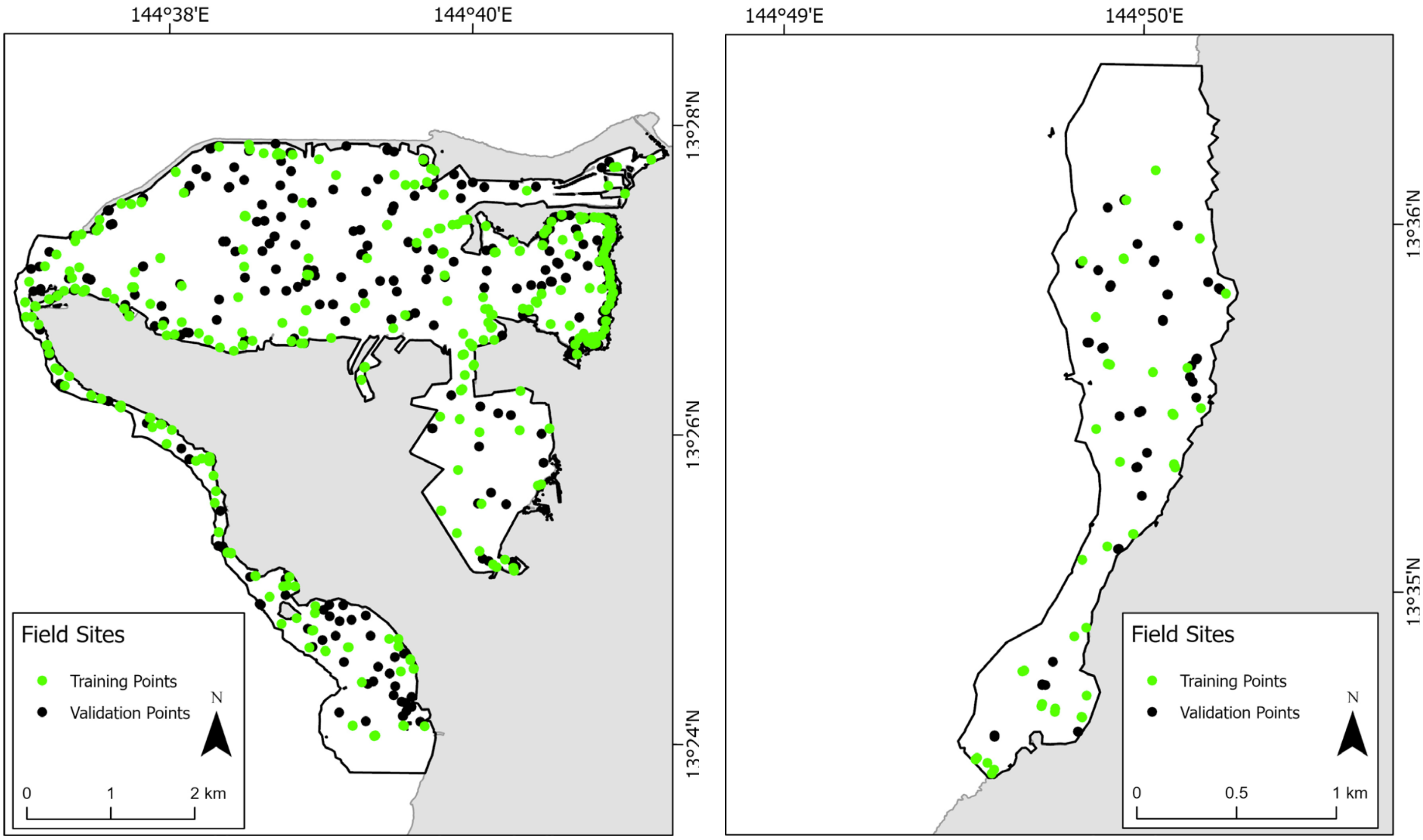


Figure 2.2. Locations where georeferenced underwater photographs were collected for model training and validation in Apra Harbor (top) and Haputo ERA (bottom).

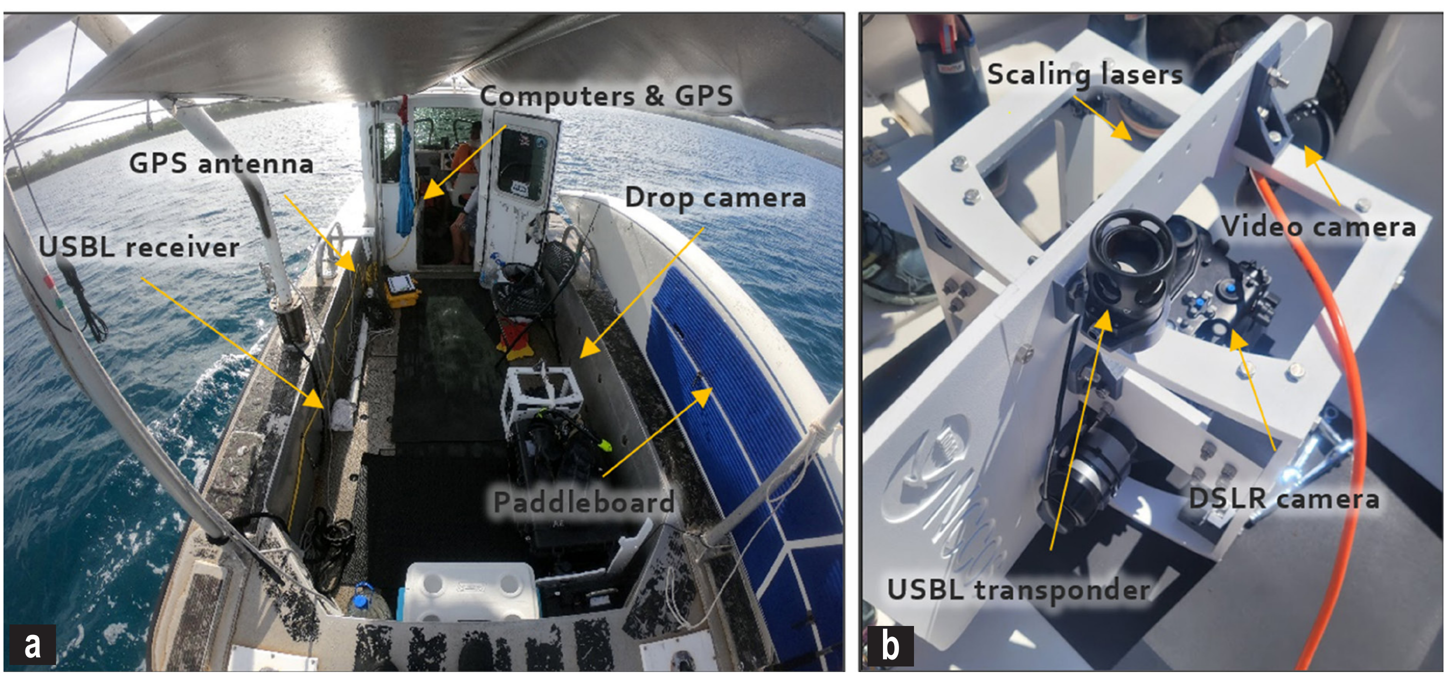


Figure 2.3. (a) Equipment used to collect training and validation data in the field. (b) The drop camera was designed and 3D printed by NOAA NCCOS and included a Sony camera (for high-resolution photographs), GoPro video camera (for real-time feed), ultra-short baseline (USBL; for underwater positioning), GPS (for above water positioning), and lasers (spaced at 10 cm for scaling). DSLR = digital single-lens reflex

2.3.2 Field Data Processing

The field data were processed to identify and annotate habitats in the photographs, and to create georeferenced photomosaics for each site. This process included four, main steps:

Underwater photographs were visually reviewed for quality, and color corrected (performed using Adobe lightroom software). The GPS data were differentially corrected using a Continuously Operating Reference Station on Guam, and reprojected to the WGS 1984 UTM 55 North coordinate system (performed using GPS Pathfinder Office software). The USBL data were exported from SeaTrac Pinpoint software, and matched with the corrected GPS locations using synchronized timestamps. Underwater photographs were georeferenced using these combined GPS-USBL locations (ESRI 2023; performed using ArcGIS Pro, Geotagged Photos To Points function). Not all images were georeferenced because sampling frequencies differed among the USBL (approximately 5 s), GPS (approximately 1 s), and cameras (approximately 0.5 s). To georeference the remaining photographs, a custom Python script was created to interpolate positions between the GPS-USBL locations. This script also mosaicked and developed 3D models from the resulting georeferenced photographs (Pierce and Winians, 2023; performed using Agisoft Metashape's application programming interface).

Seven substrate and 12 biological cover types (Table 2.1) were identified visually in the above georeferenced photographs (n = 674 annotations; training = 346; validation = 328). Multiple substrate and cover types were often present at each site. The amount of area annotated (4 m^2) was standardized in each photograph so that it matched the spatial resolution (i.e., 2 × 2 m pixels) of the environmental predictors. Percent cover was estimated to the nearest 1% for each habitat type. Percent coverages were also converted to presences (1) and absences (0) and used to train or validate the habitat predictions.

Live Acropora *coral in Apra Harbor. Credit: NOAA NCCOS*

2.4 Predicting and Classifying Benthic Habitats

BRTs and BCTs are machine learning techniques that combine regression or classification trees with boosting to model the complex, nonlinear relationships between habitat types and environmental variables. 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).

BRTs and BCTs were used for this project because they can deal with data that are 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). Please see the References and Glossary for more information.

Here, BRTs were used to develop 19 habitat predictions depicting the mean probability of occurrence for seven substrate and 12 biological cover types. Mean probabilities were calculated by creating and averaging 100 bootstrapped model iterations for each substrate and cover type (n = 1,900). BCTs were then used to create a single classified benthic habitat map using these 19 mean habitat predictions. Three main steps were used to create these map products: (1) preparing the data, (2) creating and evaluating habitat predictions, and (3) creating and evaluating a classified habitat map (Figure 2.4). This work was conducted primarily in Microsoft Azure environment using ArcGIS Pro (ESRI, 2023) and R software (R Core Team, 2022) using the dismo (Hijmans et al., 2023b), caret (Kuhn et al., 2023), and raster (Hijmans, 2023a) packages.

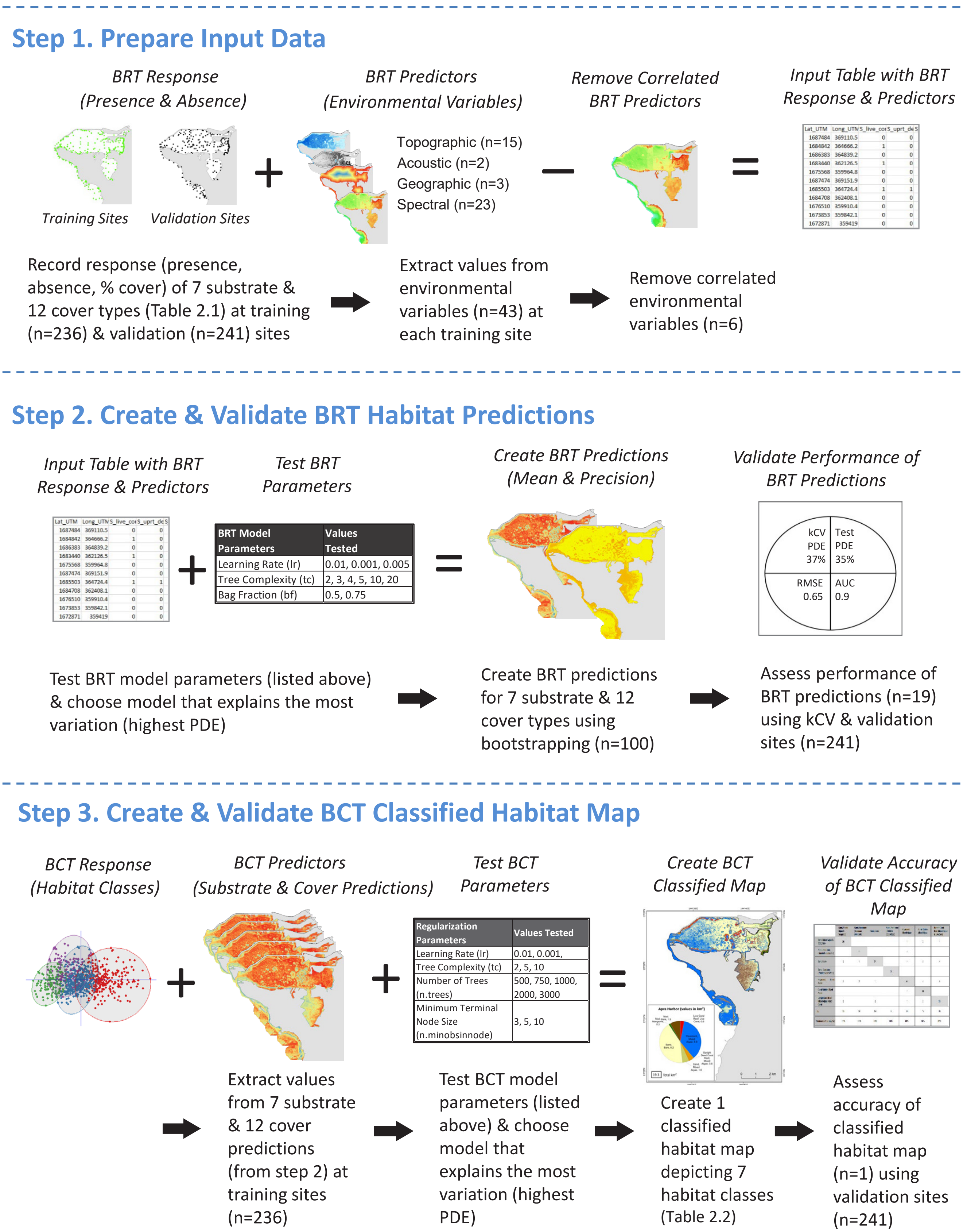


Figure 2.4. Diagram depicting steps in modeling process to predict substrate and cover distributions and develop a classified benthic habitat map. BRT = boosted regression tree; BCT = boosted classification tree; PDE = percent deviance explained; kCV = k-fold cross validation; RMSE = root mean square error; AUC = area under the curve

2.4.1 Step 1 – Prepare Input Data

The presence (1) or absence (0) of seven substrate and 12 biological cover types was used 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 43 environmental (i.e., spectral, geographic, topographic) predictors were numeric. For the spectral predictors, natural color satellite imagery was used to predict mangrove habitats above the water, and water column–corrected satellite imagery was used to predict marine habitats below the water. These predictors differed because correcting the imagery for water column changes was not relevant to mangrove habitats.

Pairwise testing was conducted to identify and remove predictors that were highly correlated (i.e., Spearman rank $\rho \geq 0.9$ or $\rho \leq -0.9$) with three or more other predictors. One predictor (water column corrected satellite band 12 Red Coastal Blue) was removed from the predictor set for marine habitats. Five predictors (natural color satellite spectral bands 1–5) were highly correlated and removed from the predictor set for mangroves. The training sites (i.e., locations denoting the presence or absence of substrate and cover types) were intersected with the remaining environmental predictors to extract their value at each location. This spatial intersection combined the response and predictor datasets into a single table used in step 2.

2.4.2 Step 2 – Create and Validate BRT Habitat Predictions

In this step, the table with the response and predictor values was used to test different BRT model tuning parameter combinations in R (R Core Team, 2022; Hijmans et al., 2023b; performed using the dismo package). A range of input values were tested for the following tuning parameters: 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. The bf specifies the proportion of data that is randomly chosen at each step. The larger the bf, the more data available to train the model at each step. For each of the seven substrate and 12 cover types, 36 combinations of lr, tc, and bf were tested (Table 2.3). k-fold cross validation (kCV) was used to identify the combinations of lr, tc, and bf that created the model that explained the most variation. 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.

This process was repeated 10 times (i.e., one time for each fold) × 36 model parameter combinations × 19 substrate and cover types (n = 6,840 models total). Model performance was measured using PDE, which was calculated using the one remaining fold and then averaged across the 10 folds. PDE is the amount (%) of variation explained in the response data. PDE values normally range between 0% and 100%, with higher values indicating better model performance. The models with the highest PDE were selected for each substrate and cover type (n = 19 models). The remaining models were discarded.

These 19 best models were then applied spatially to predict the distribution of the seven substrate and 12 cover types throughout the project area (R Core Team, 2022; Hijmans, 2023a; performed using the raster package in R). These raster predictions represent the average of 100 model iterations created using bootstrapping (See Glossary for terminology) for each substrate and cover type. These predictions describe the mean probability of occurrence for each habitat (i.e., the likelihood [%] that a particular substrate or cover type is present in a pixel). Larger probabilities indicate it is more likely the substrate or biological cover type is present.

The precision associated with each probability of occurrence prediction was also quantified using the same 100 bootstrapped model iterations for each substrate and cover type. Precision is reported as the coefficient of variation (CoV), which represents the standard deviation as a proportion of the mean. Instead of reporting two values (i.e., minimum and maximum), CoV captures the range of probabilities in a single value for each pixel. CoV can be multiplied by the probability of occurrence to derive the standard deviation and thereby quantiles and confidence intervals for the estimated probabilities in a pixel. Smaller CoVs indicate that the prediction has higher precision and less uncertainty. Larger CoVs indicate there is more uncertainty associated with the spatial prediction. Sometimes, large CoVs occur artificially because the mean predicted values are extremely small (most notably in areas of predicted absence). Consequently, CoVs should be viewed along with the mean prediction to avoid misinterpretation.

Table 2.3. Suite of boosted regression tree (BRT) model parameters and values tested.

Regularization Parameters	Parameters Tested	Definition
Learning Rate (lr)	0.01, 0.001, 0.005	Determines contribution of each tree to the growing model
Tree Complexity (tc)	2, 3, 4, 5, 10, 20	Controls how many predictor interactions are fitted in a tree
Bag Fraction (bf)	0.5, 0.75	Controls proportion of data randomly selected to build each tree

Finally, the performance of the substrate and biological cover predictions was evaluated using five different metrics: (1) kCV PDE, (2) test PDE, (3) test bias, (4) test RMSE, and (5) test area under the ROC curve (AUC). kCV PDE was calculated during kCV 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, test bias, test RMSE, and test AUC were independently calculated using the validation sites. Test PDE, like kCV PDE, is the amount (%) of variation explained in the response data. PDE values normally range between 0% and 100%, with higher values indicating better model performance. Conversely, bias and RMSE measure 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 photographs). Here, bias is used to describe the direction (+ or −) of the error, and RMSE is used to describe the size of the error. Bias closer to 0 and lower RMSE denote better model performance.

ROC curves measure a model's predictive performance in a different way compared to PDE, bias, 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." 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. Spatial autocorrelation of model residuals was also tested using global Moran's I (R Core Team, 2022; performed using the ape package in R). Five different metrics (plus Moran's I) were calculated because they describe model performance in different ways and, when viewed together, can provide a more thorough understanding of the model's limitations.

2.4.3 Step 3 – Create and Validate BCT Classified Habitat Map

In this step, BCTs were used to develop a classified habitat map depicting the distribution of the seven habitats identified by cluster analysis (Table 2.2). This response variable was modeled using a multinomial (many groups) distribution. The 19 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 $\rho < 0.9$ or $\rho > -0.9$). The 236 training sites (each of which were assigned one of the seven habitat types) were intersected with the 19 probability of occurrence predictions to extract their value at each location. This spatial intersection combined the training and predictor values into a single table.

Next, BCT models were fit and tuning parameters optimized in R (R Core Team, 2022; Kuhn et al., 2008; performed using the caret package). One hundred and eighty combinations of lr, tc, number of trees (n.trees), and minimum terminal node size (n.minobsinnode) were tested (Table 2.4). 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 modeling 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. kCV PDE was used to identify the combinations of lr, tc, n.trees, and n.minobsinnode that created the highest performing model. This highest performing model was then applied spatially to create the classified habitat map for the project areas.

Table 2.4. Suite of boosted classification tree (BCT) model parameters and values tested.

Regularization Parameters	Parameters Tested	Definition	Impact	Definition
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 classified habitat map was qualitatively assessed by local experts on Guam (Appendix B), and quantitatively assessed using 328 photographs at 241 validation sites. The validation sites were grouped into the same seven habitats identified by the cluster analysis. Sites were considered correct if the same habitat was present within 6.1 m (approximately 3 pixels) of the validation site due to the ±6.1-m positional uncertainty of the satellite mosaic. A confusion matrix was developed using the validation data describing the classified map's overall accuracy, producer's accuracy (PA), and user's accuracy (UA; Story and Congalton, 1986). This matrix was constructed as an array with seven rows (denoting the predicted classification by the BCT) and seven columns (denoting the observed classification from validation sites). The overall accuracy was calculated as the sum of the major diagonal (i.e., matching predicted and observed classifications), divided by the total number of validation 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 classified 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 PA, and by the row total (n_j) to yield a UA. The tau coefficient was also calculated to account for the random, chance agreement between the map and training 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).

Soft and hard live corals near Spanish Steps. Credit: NOAA NCCOS

Live Acropora *coral near Spanish Steps. Credit: NOAA NCCOS*

*Sponges (*Ianthella basta*) and* Halimeda *algae in Apra Harbor. Credit: NOAA NCCOS*

Sea star in Apra Harbor. Credit: NOAA NCCOS

Chapter 3 Results and Discussion

Free diving to explore benthic habitats in Haputo ERA. Credit: NOAA NCCOS

Over 21 km^2 of seafloor was characterized in and around Apra Harbor (19 km^2) and Haputo ERA (2 km^2) from 0- to approximately 50-m depths. This section presents the results from these models, highlights some of the main features of the habitat predictions, reports the performance of the habitat predictions and accuracy of the maps, and discusses the limitations and potential applications of these products to meet particular research and management needs.

3.1 Model Performance

Nineteen BRT models and resulting spatial predictions describe the probabilities of occurrence for seven substrate and 12 cover types. Prevalence (i.e., number of presences divided by the total number of samples) of these habitats ranged from common (e.g., 74% for turf algae) to rare (e.g., 1% for *Halodule uninervis* seagrass). Despite these differences, model performance was considered "good to excellent" based on five evaluation metrics. Specifically, kCV PDE ranged from 14.9% to 88.7% ($\bar{x}$ = 39.3% ±4.1 SE), and test PDE ranged from 3.8% to 86.7% ($\bar{x}$ = 27.3% ±5.0 SE). The Mangrove model had the highest kCV and test PDEs (88.7% and 86.7%). The Rubble model had the lowest kCV PDE (14.9%), and the Live Coral (Foliose) model had the lowest test PDE (3.8%). Test AUC values ranged from 0.70 (good) to 0.99 (excellent) for all the models ($\bar{x}$ = 0.86 ±0.02 SE). The Mangrove model had the highest test AUC (0.99), and Rubble and Live Coral (Foliose) models had the lowest test AUC (0.70).

Test bias was small for all models, ranging between −0.1 to +0.05 ($\bar{x}$ = -0.02 ±0.01 SE). Bias indicates whether the model under predicted (−) or over predicted (+) the probability of occurrence. The Seagrass (*Halodule uninervis*), Live Coral (*Porites rus*), and Live Coral (Encrusting) models showed no systematic test bias. The Mud, Sponge, and Mangrove models had a positive test bias, and consistently over-predicted probabilities by 0.05, 0.04, and 0.01, respectively. The remaining models had negative test biases, and under-predicted probabilities by 0.01 to 0.1. Lastly, test RMSE values ranged from 0.08 to 0.47 ($\bar{x}$ = 0.32 ±0.03 SE). The Sponge model had the largest amount of error (0.47), while the Seagrass (*Halodule uninervis*) had the smallest (0.08).

Spatial autocorrelations of model residuals were also quantified using Moran's I. Residuals were autocorrelated if the probability value (p value) was ≤ 0.05, indicating the observed value of I was significantly different from expected value. Here, probabilities ranged from 0 to 0.53 for all models. Residuals for four models (Live Coral (Branching), Live Coral (Encrusting), Bare and Rock) were not spatially autocorrelated ($p \geq 0.15$). Residuals for the remaining 15 of the 19 models were spatially autocorrelated ($p \leq 0.03$). This pattern suggests that there are influential environmental predictors missing from the BRT modeling process for 15 of the 19 habitat predictions.

3.2 Geographic patterns of substrate and cover types

Substrate: Live Coral (All Species)

"Live Coral (All Species)" (Figure 3.1a) was common and present at 49% (236/480) of the training and validation sites (Figure 3.1b). Most observations were documented between Point Udall and Apaca Point or clustered along the reef crests and patch reefs inside Outer Apra Harbor. Live coral was also found through Haputo ERA. The Live Coral (All Species) model showed a similar spatial pattern, with the highest likelihood of presence along the shoreline from San Luis to Acapa Points, from Tristar Dock to the mouth of Outer Apra Harbor and at reefs west and east of Dry Dock Island (Figure 3.1c). Probability of occurrence values were also high throughout Haputo ERA. The maximum probability was 98% for Live coral (All Species). CoV values were lowest (<0.25) in these same locations (Figure 3.1d), 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 the "Coral" habitat class depicted in the 2005 map by NCCOS (NOAA NCCOS, 2005). The one notable exception is the area from Point Udall to Tantapalo Point, where live coral was predicted in 2023 but not mapped in 2005. This exception is due to the different scales used in the 2005 map (4,047 m^2) versus the 2023 map (4 m^2).

Substrate: Upright Dead Coral Reef

"Upright Dead Coral Reef" (Figure 3.2a) was fairly common and was present at 37% (175/480) of the training and validation sites (Figure 3.2b). Upright dead coral reef was primarily present along the reef crest from San Luis Beach to Orote Island, offshore Dadi Beach, along the Glass Breakwater, and at reefs west of Dry Dock Island and in Sasa Bay. Upright dead coral reef was also found through Haputo ERA. The model showed similar spatial patterns, with the highest likelihood of presence in these same locations (Figure 3.2c). Probability of occurrence values were also moderately high throughout Haputo ERA. The maximum probability was 84% for dead reef. CoV values were lowest (<0.25) in these same locations (Figure 3.2d), 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," "Individual Patch Reef," and "Spur and Groove" in the 2005 NCCOS map (NOAA NCCOS, 2005), except in Sasa Bay which was primarily mapped as "Pavement."

Substrate: Pavement

"Pavement" was common in the project area and was present at 51% (246/480) of the training and validation sites (Figure 3.3b). "Pavement" was concentrated from San Luis Beach around Point Udall to Acapa Point. Pavement was also dispersed on reef features throughout Outer Apra Harbor and Haputo ERA. The Pavement model showed similar spatial patterns, with the highest likelihood of presence on reef crests around Orote Peninsula (Figure 3.3c). Probability of occurrence values were also moderately high throughout Haputo ERA. The maximum probability was 94% for "Pavement." CoV values were lowest (<0.25) in these same locations (Figure 3.3d), indicating higher precision and lower uncertainty for places where pavement is more likely to be present. These spatial patterns for "Pavement" match the 2005 NCCOS map (NOAA NCCOS, 2005) in Haputo ERA and around Orote Peninsula, but they differ in Sasa Bay and along the north side of Outer Apra Harbor.

Substrate: Rock

"Rock" (Figure 3.4a) was present at 9% (43/480) of the training sites and was rare except for sites along the coastline from Orote Island to just north of Apuntua Point, and along the Glass Breakwater, which was built from limestone post World War II (Figure 3.4b). Rock was not present in Haputo ERA. The Rock model showed similar spatial patterns, with the highest likelihood nearshore the Glass Breakwater and southern Orote Peninsula (Figure 3.4c). Probabilities were moderate to low in Haputo. The maximum probability was 93% for "Rock." Like the other models, CoV values were lowest (<0.5) in these same locations (Figure 3.4d), indicating higher precision and lower uncertainty. However, CoV was high (>1) everywhere else in the project area. "Rock" was not mapped in the 2005 NCCOS map (NOAA NCCOS, 2005) in Apra Harbor or Haputo ERA.

Substrate: Rubble

"Rubble" (Figure 3.5a) was present at 27% (129/480) of the training sites (Figure 3.5b). It was more common in Apra Harbor and less common in Haputo ERA. Overall, "Rubble" distributions were similar to "Pavement" and were primarily present offshore from San Luis Beach to Dadi Beach, along the Glass Breakwater's southern shoreline, south of Cabras Island, and at reef features in Outer Apra Harbor. This habitat was also present in the southern third of Haputo ERA. The Rubble model showed similar spatial patterns with higher probabilities offshore Dadi Beach, and moderate to low probabilities also being predicted in the center of Outer Apra Harbor and in Haputo (Figure 3.5c). The maximum probability was 81% for "Rubble." Similar to the Rock model, CoV values were lowest (<0.25) in locations with high probabilities (Figure 3.5d) but moderate to high (>0.75) in all other locations. In both the CoV and mean prediction surfaces, there is also a visual artifact starting near San Luis Beach trending north to the Glass Breakwater from a seamline in the satellite image mosaic. Compared to the 2005 NCCOS map, "Rubble" was predicted over a much larger geographic area (NOAA NCCOS, 2005).

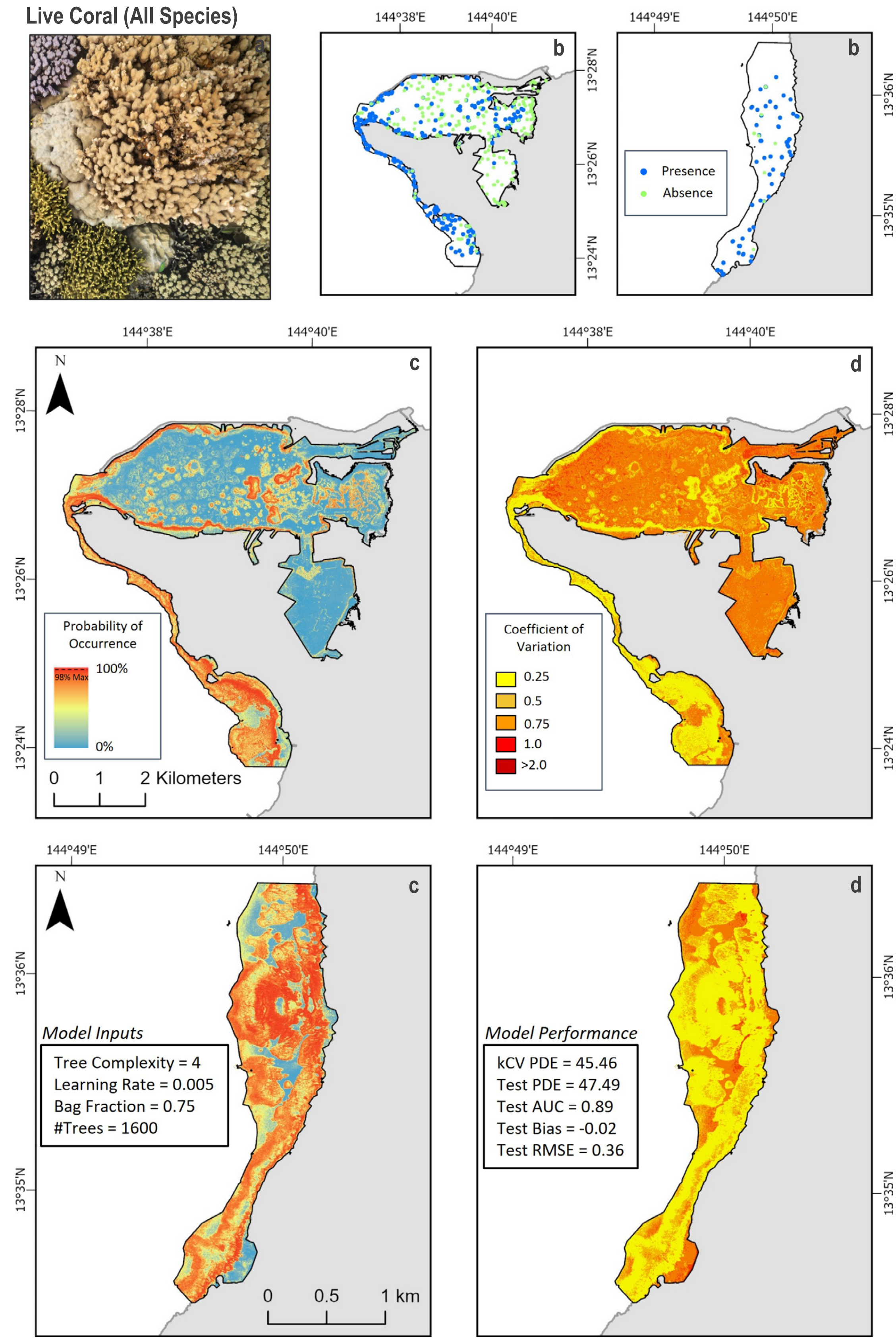


Figure 3.1. Predicted presence of "Live Coral (All Species)." Figure panels depict: a) a photo of live coral reef; b) maps denoting the presences and absences of live coral in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

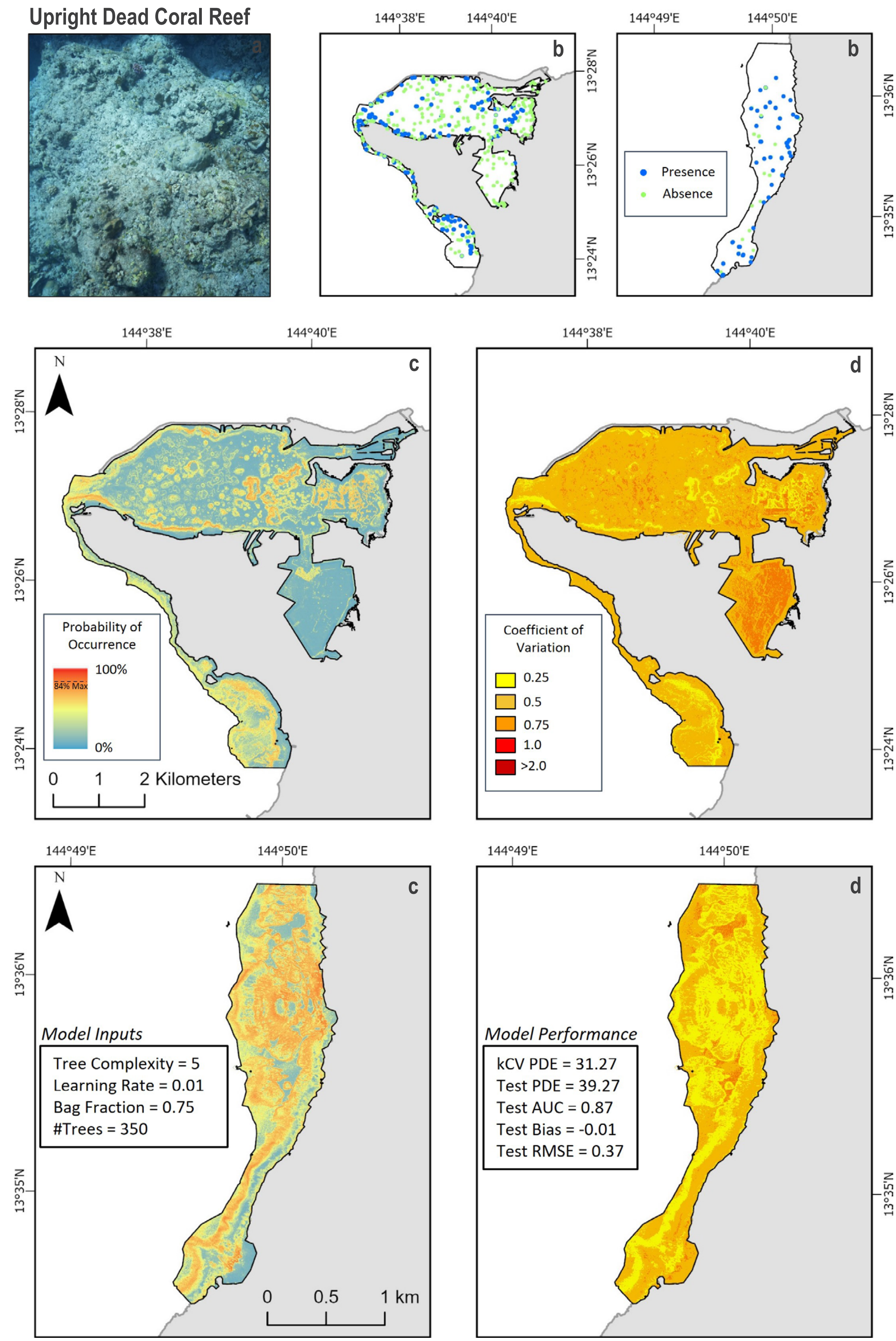


Figure 3.2. Predicted presence of "Upright Dead Coral Reef." Figure panels depict: a) a photo of upright dead coral reef; b) maps denoting the presences and absences of upright dead coral reef in the and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

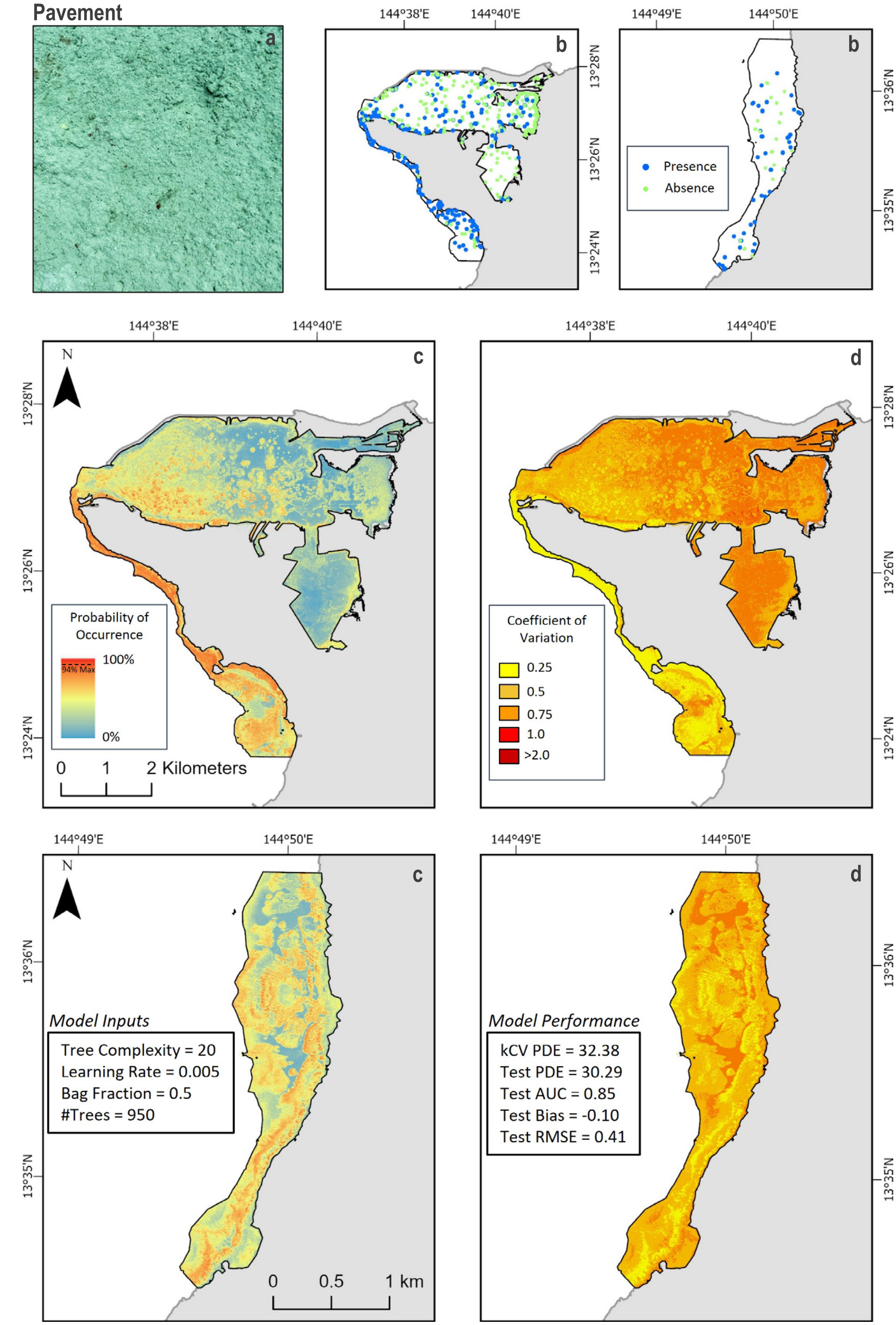


Figure 3.3. Predicted presence of "Pavement." Figure panels depict: a) a photo of pavement; b) maps denoting the presences and absences of pavement in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

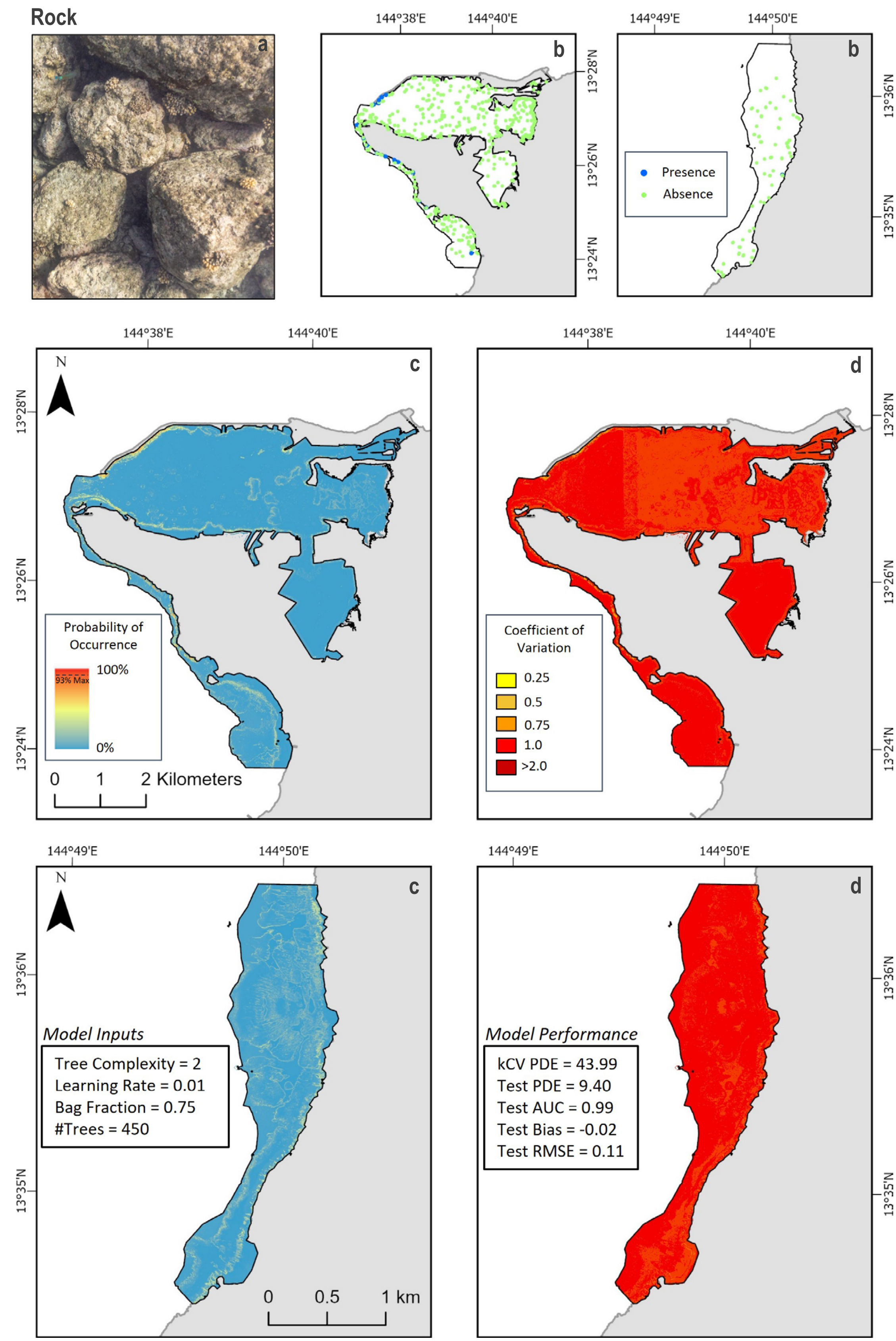


Figure 3.4. Predicted presence of "Rock." Figure panels depict: a) a photo of rock; b) maps denoting the presences and absences of rock in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

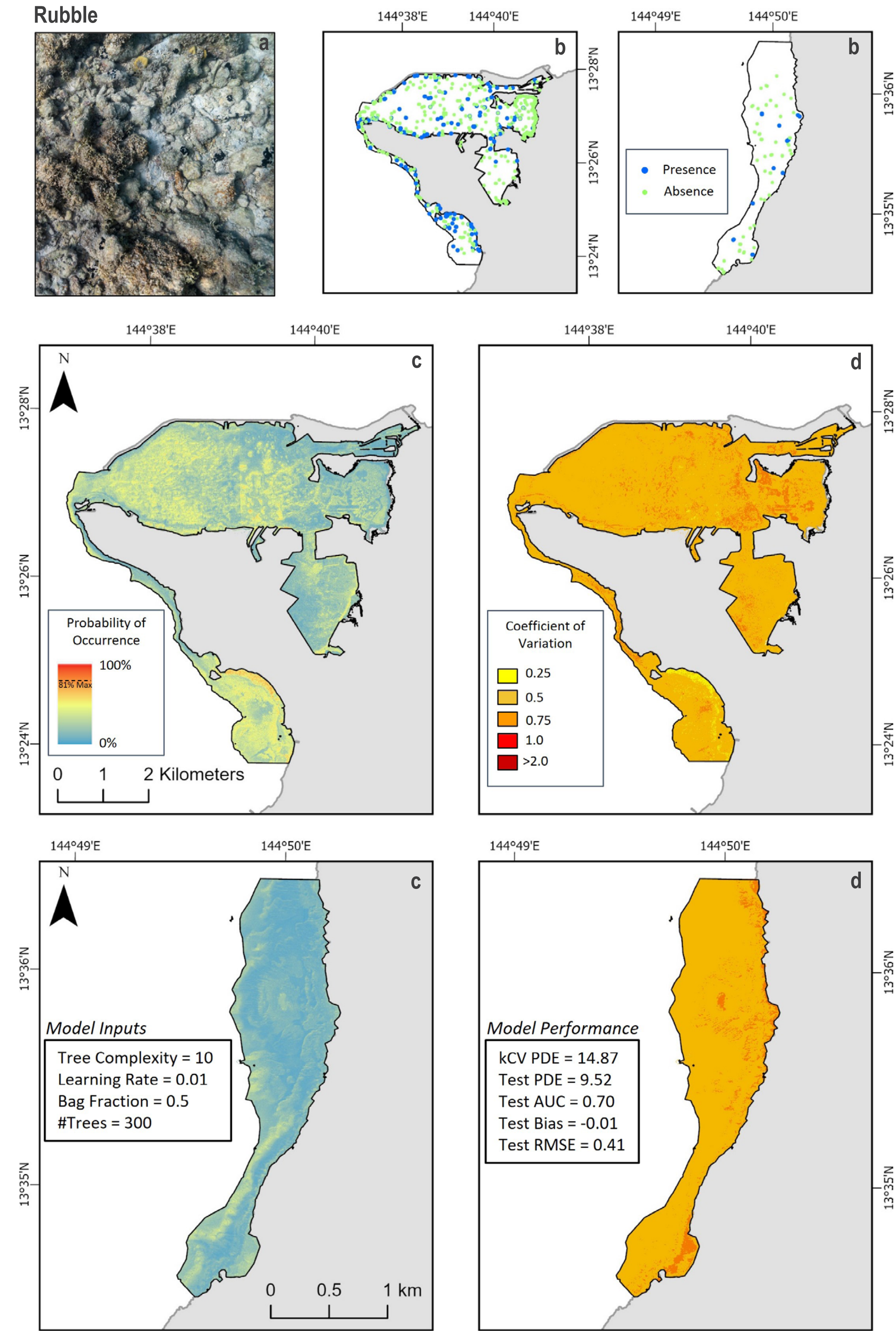


Figure 3.5. Predicted presence of "Rubble." Figure panels depict: a) a photo of rubble; b) maps denoting the presences and absences of rubble in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

Substrate: Sand

"Sand" (Figure 3.6a) was present at 57% (272/480) of the training sites and was most common in Outer Apra Harbor, offshore Dadi Beach and Acapa Point, as well as in deeper areas inside Haputo ERA (Figure 3.6b). However, it had notably lower abundances inside Inner Apra Harbor, from Point Udall to Tipalao Bay and in the northern section of Haputo ERA. The Sand model showed similar spatial patterns, with the highest likelihood of sand in the center of Outer Apra Harbor, Sasa Bay, south of Cabras Island, and in depths greater than 25 m offshore Dadi Beach, and greater than 15 m in Haputo ERA (Figure 3.6c). The maximum probability was 96% for "Sand." Like the other models, CoV values were lowest (<0.25) in these same locations (Figure 3.6d), 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 2005 NCCOS map (NOAA NCCOS, 2005). The notable exception is Inner Apra Harbor, which was mapped as "Sand" in 2005 but predicted to be both "Mud" and "Sand" in 2023.

Substrate: Mud

"Mud" (Figure 3.7a) was present at 11% (53/480) of the training sites and was most common inside Inner Apra Harbor and near the shoreline in Sasa Bay (Figure 3.7b). It was not present in Haputo ERA. The Mud model showed similar spatial patterns, with the highest likelihood in Inner Apra Harbor and around the mangroves in Sasa Bay (Figure 3.7c). The maximum probability was 90% for "Mud." Like the other models, CoV values were lowest (<0.5) in these same locations (Figure 3.7d) but high (>1) everywhere else. These spatial patterns broadly match the distributions of the "Mud" class in the 2005 NCCOS map (NOAA NCCOS, 2005). The notable exception is Inner Apra Harbor, which was mapped as "Sand" in 2005 but predicted to be both "Mud" and "Sand" in 2023.

Cover: Mangrove

Mangroves (Figure 3.8a) were very rare in the project areas, occurring at only 3% (15/480) of the training and validation sites (Figure 3.8b). No mangroves were documented and modeled in Haputo ERA. All of the mangrove occurrences were inside the Sasa Bay and Inner Apra Harbor, and absent everywhere else in the project area. The Mangrove model showed similar spatial patterns, with the highest likelihood of mangrove presence along the Sasa Bay shoreline and at the Abo Cove and Atantano River in Inner Apra Harbor (Figure 3.8c). The maximum probability was 99% for "Mangrove." Unlike the previous models, the CoV values were low (<0.25) where the model predicted high and low probabilities of occurrence (Figure 3.8d). CoV was high (>1) in near the shoreline around Orote Peninsula and Glass Breakwater, and further inland in Sasa Bay and Inner Apra Harbor. This "Mangrove" probability of occurrence prediction broadly matches the distributions of the "Emergent Vegetation" class in the 2005 NCCOS map (NOAA NCCOS, 2005). The notable exception is in Inner Apra Harbor, where the 2005 NCCOS map did not depict any "Emergent Vegetation."

Cover: Live Coral (Branching Corals)

"Live Coral (Branching Coral)" (Figure 3.9a) occurred at 19% (93/480) of the training and validation sites (Figure 3.9b). Inside the Harbor, branching corals were present on or near prominent dive locations, including Dogleg Reef, Dry Dock Reef, Seabee Junkyard, Fingers Reef and Gab Reef. Outside the Harbor, they primarily occurred around Orote Island, from Point Udall to Acapa Point and in the southern three-quarters of Haputo ERA. The branching coral model showed similar spatial patterns, with the highest likelihood of these taxa being present around Orote Island and in less than two m depths inside Haputo ERA (Figure 3.9c). The maximum probability was 85% for branching corals. Like the other models, CoV values were lowest (<0.25) in these same locations (Figure 3.9d), indicating higher precision and lower uncertainty for predicted presences. CoV were moderate (>0.75) everywhere else. No comparison was made to the 2005 NCCOS map (NOAA NCCOS, 2005) because this taxonomic group was not explicitly mapped in that study.

Cover: Live Coral (Encrusting)

"Live Coral (Encrusting)" (Figure 3.10a) were present at 36% (173/480) of the training and validation sites (Figure 3.10b), and was most common near the mouth of Apra Harbor, from Point Udall to Acapa Point and throughout Haputo ERA. The remaining presences were sparsely west of Dry Dock Island and in Sasa Bay. The Live Coral (Encrusting) model showed similar spatial patterns, with moderate probabilities from the mouth of Apra Harbor to Acapa Point (Figure 3.10c). A few other areas also had moderate to high probabilities of occurrence, including the reefs in the northern section of Haputo ERA. The maximum probability was 74% for encrusting corals. Like other habitat models, the CoV values were lowest (<0.5) where the model predicted the highest probabilities of occurrence and highest (>0.75) where the model predicted low probability of occurrence values (Figure 3.10d). Like with branching corals, no comparison was made to the 2005 NCCOS map (NOAA NCCOS, 2005) because this taxonomic group was not explicitly mapped in that study.

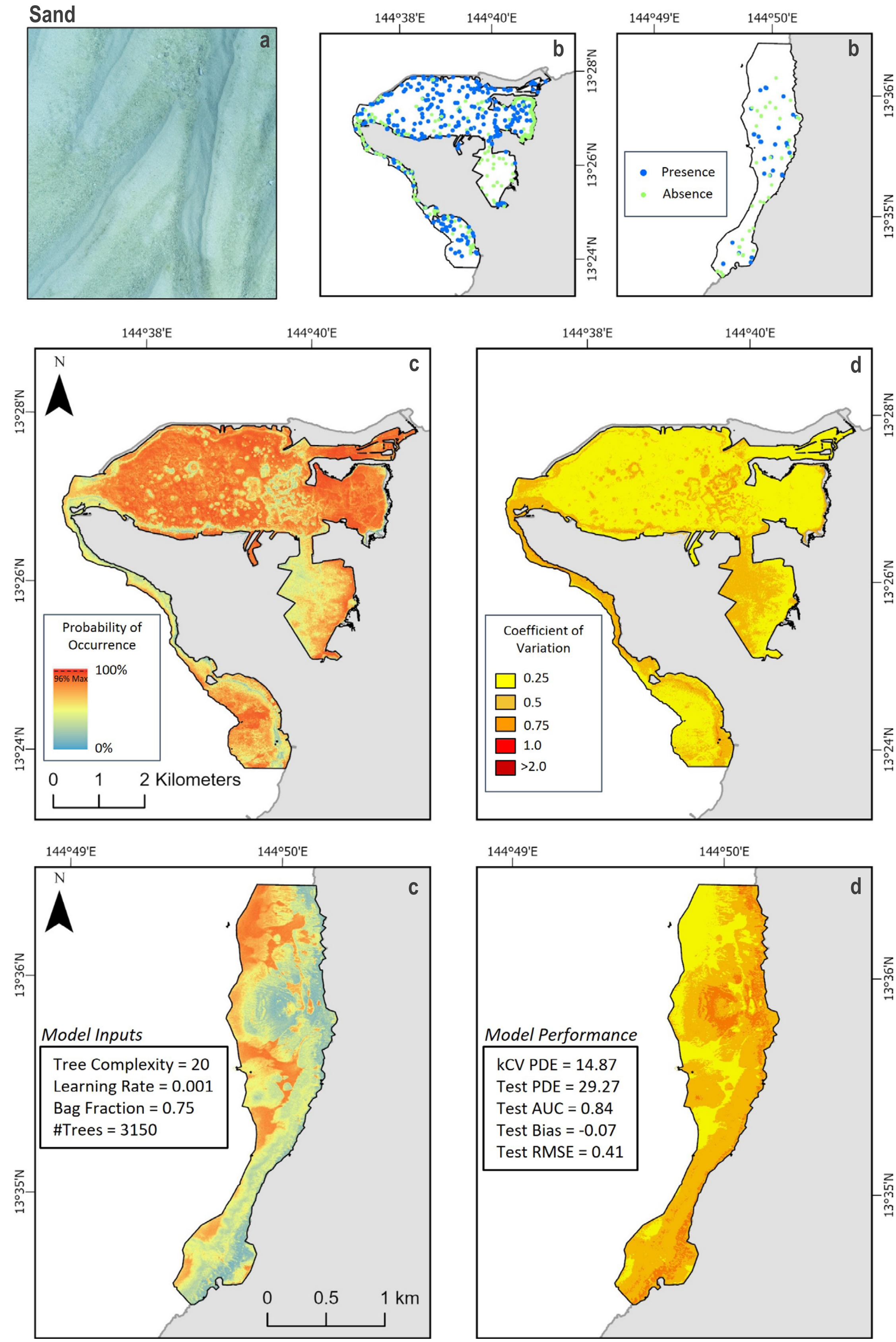


Figure 3.6. Predicted presence of “Sand.” Figure panels depict: a) a photo of sand; b) maps denoting the presences and absences of sand in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

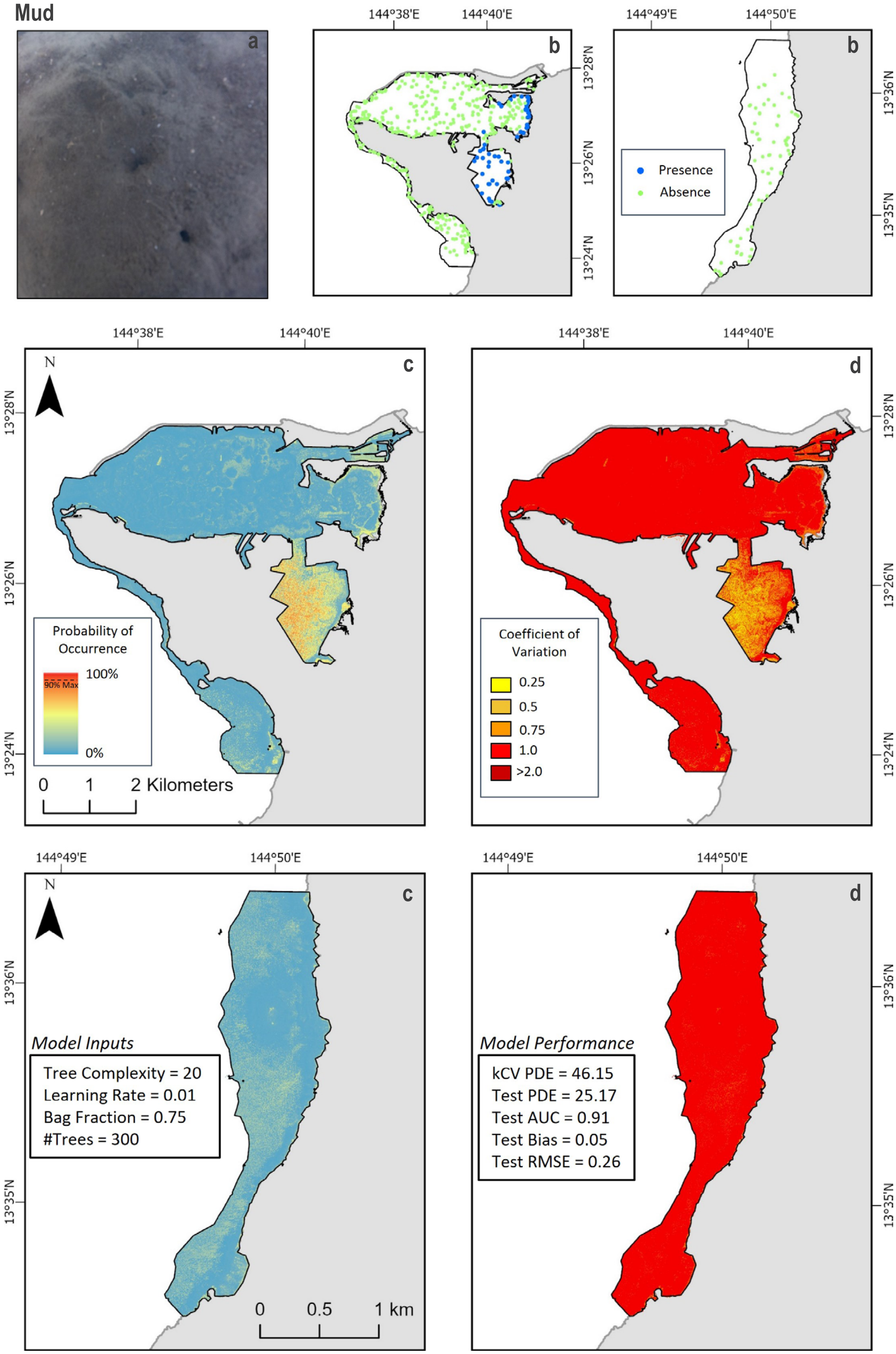


Figure 3.7. Predicted presence of "Mud." Figure panels depict: a) a photo of mud; b) maps denoting the presences and absences of mud in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

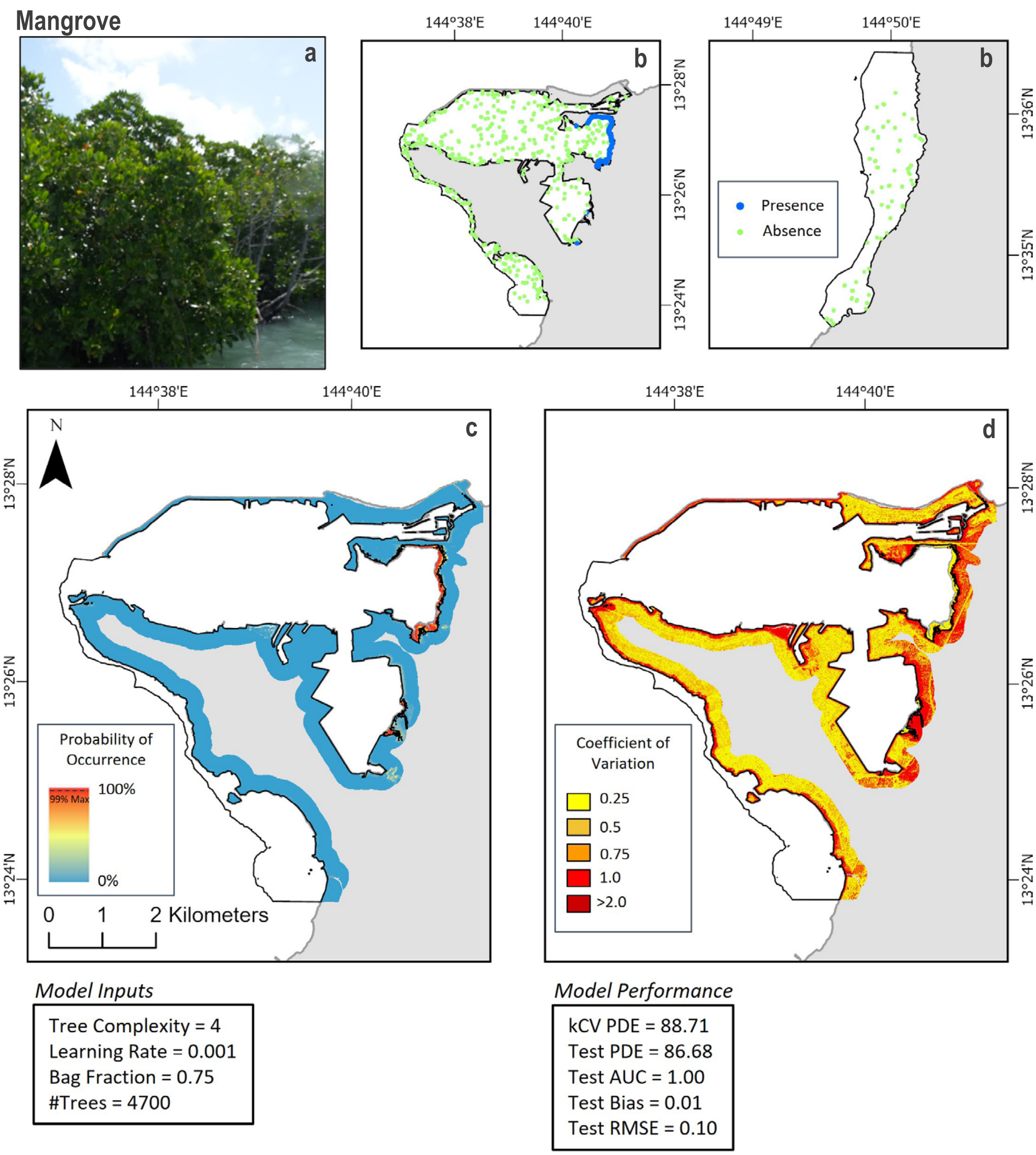


Figure 3.8. Predicted presence of "Mangrove." Figure panels depict: a) a photo of mangrove; b) maps denoting the presences and absences of mangrove in the training and validation data; c) a map denoting the predicted average probability of occurrence; and d) a map denoting the coefficient of variation. The insets at the bottom show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error..

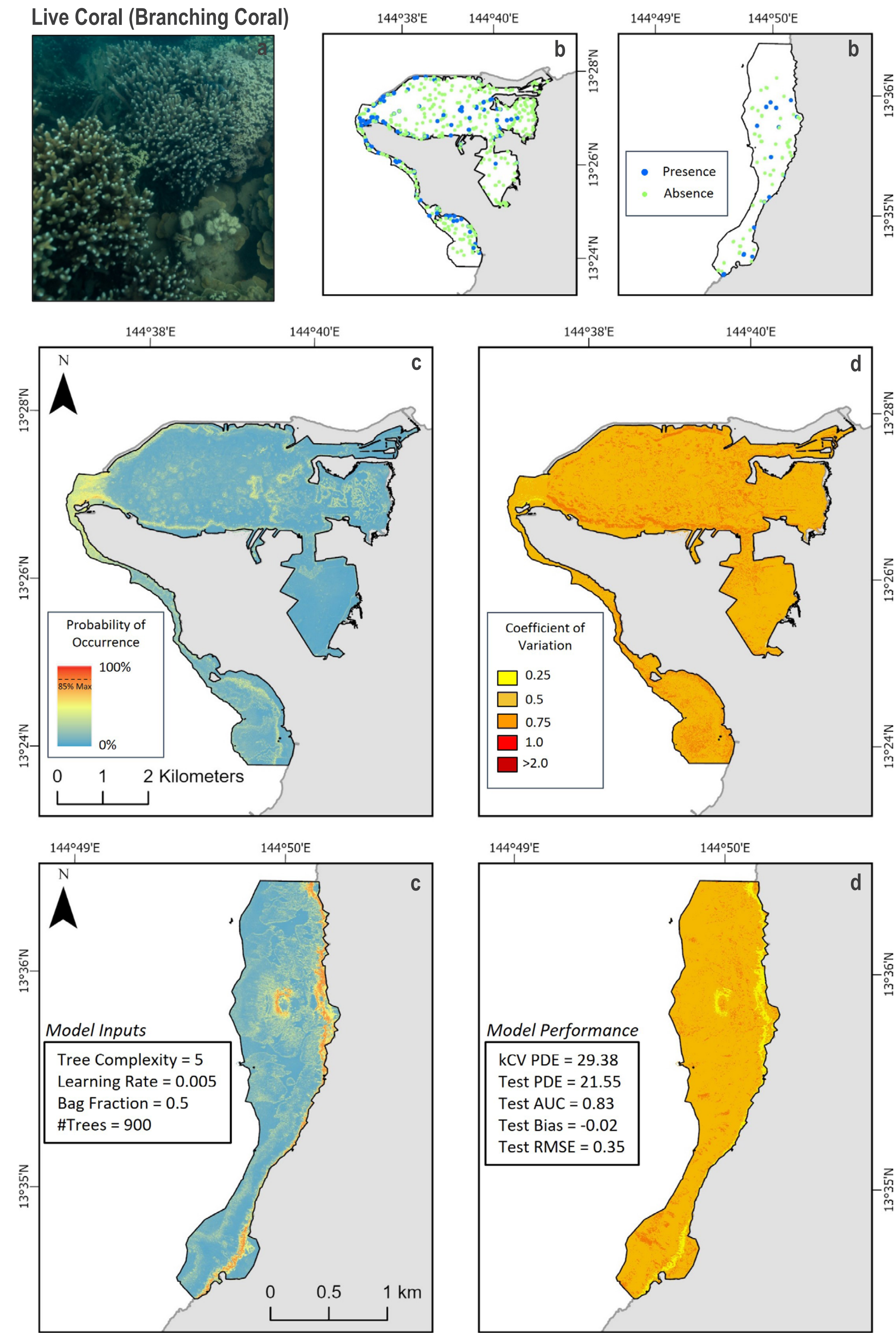


Figure 3.9. Predicted presence of "Live Coral, Branching Coral." Figure panels depict: a) a photo of branching coral species; b) maps denoting the presences and absences of branching corals in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

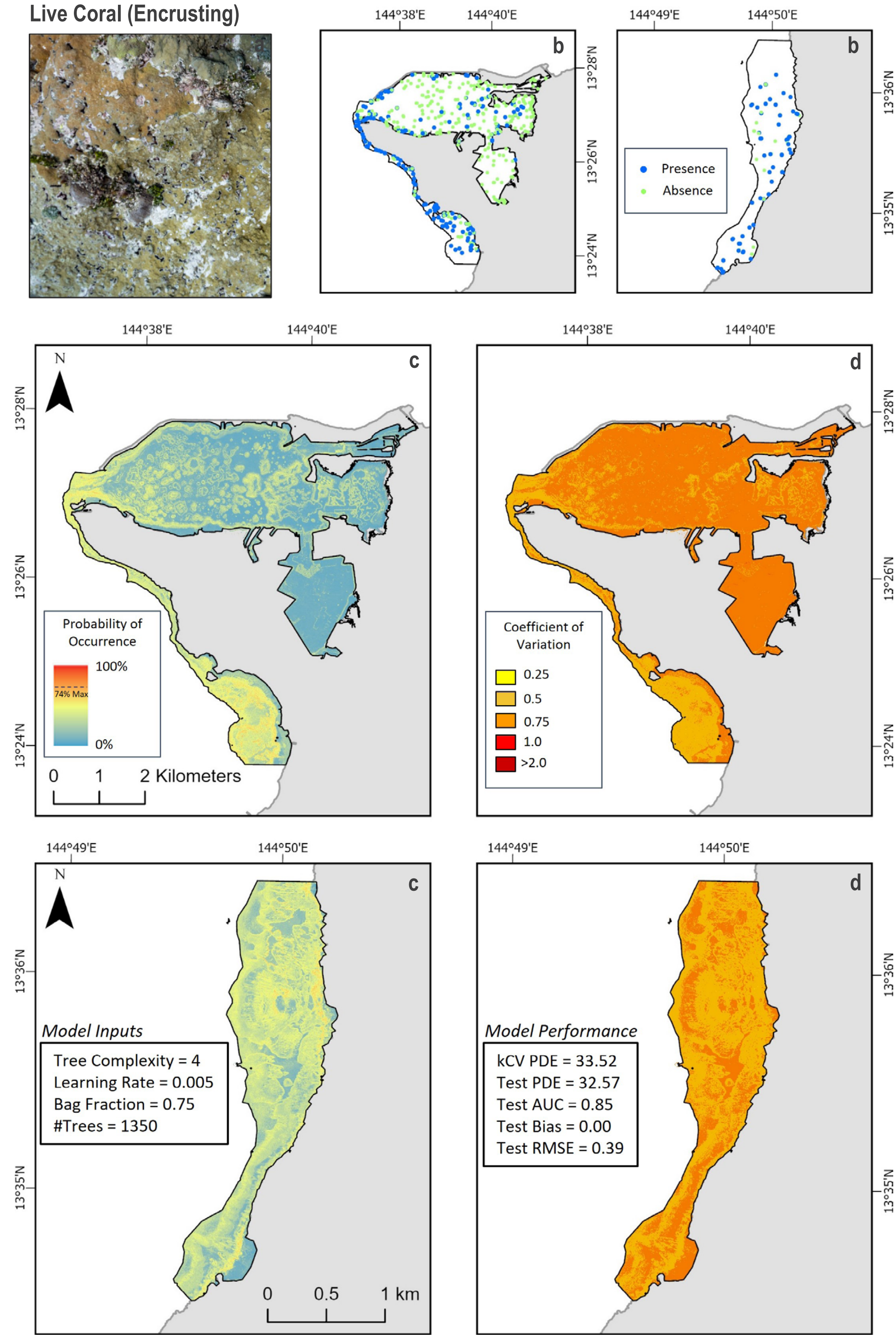


Figure 3.10. Predicted presence of "Live Coral (Encrusting)." Figure panels depict: a) a photo of encrusting corals; b) maps denoting their presences and absences of encrusting corals in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

Cover: Live Coral (Foliose)

"Live Coral (Foliose)" (Figure 3.11a) were rare and present at 5% (25/480) of the training sites (Figure 3.11b). This taxonomic group was most common in depths greater than 30 m inside Outer Apra Harbor. Only two sites had foliose corals inside Haputo ERA. The Live Coral (Foliose) model showed similar spatial patterns, with the relatively higher likelihood of presences on the sides of reefs west of Dry Dock Island, and the fore reef from San Luis Beach to Orote Island (Figure 3.11c). Higher probabilities of occurrence were also found on fore reefs offshore Haputo Beach and in the northern half of Haputo ERA. However, the maximum probability was 33% for foliose corals, and these higher probabilities were still very low compared to other taxonomic groups. The CoV values were lowest (<0.5) where the model predicted the lowest probabilities of occurrence in both Apra Harbor and Haputo ERA (Figure 3.11d). Like with branching and encrusting corals, no comparison was made to the 2005 NCCOS map (NOAA NCCOS, 2005) because this taxonomic group was not explicitly mapped in that study.

Cover: Live Coral (*Porites rus*)

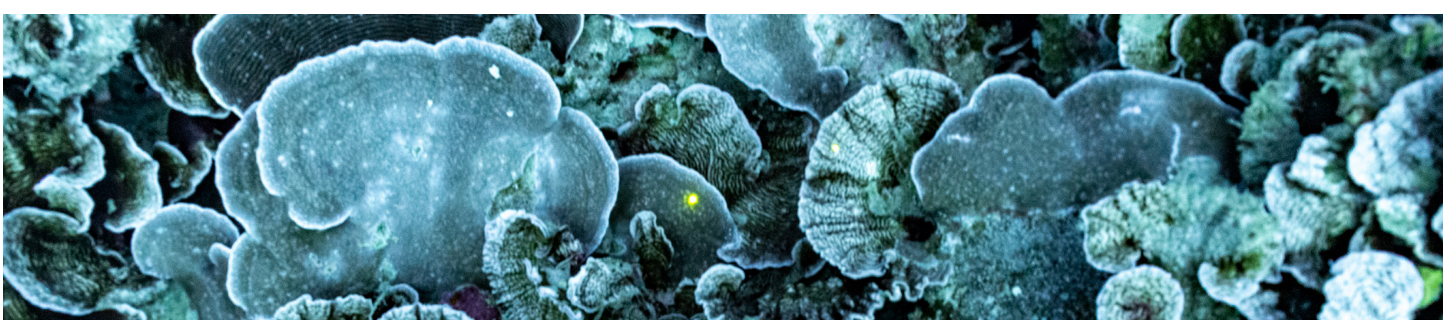
"Live Coral (*Porites rus*)" (Figure 3.12a) was present at 19% (90/480) of the training sites (Figure 3.12b). This species was most common along fore reefs in Outer Apra Harbor and throughout Haputo ERA. The remaining presences were concentrated offshore Dadi Beach and Acapa Point. The Live Coral (*Porites rus*) model showed similar spatial patterns, with the highest likelihood of presence along fore reefs in Outer Apra Harbor, at reefs west Dry Dock Island, inside Sasa Bay and along the fore reef offshore Dadi Beach (Figure 3.12c). An expert reviewer commented that *Porites rus* probabilities should have been higher on deeper reefs (>15 m) offshore Gab Gab Beach (Appendix B). Higher probabilities of occurrence were also predicted on the shelf and fore reefs in Haputo ERA. CoV values were lowest (<0.25) where the model predicted the highest probabilities of occurrence. Pockets of high CoV values (>1) coincided with low probabilities, and were dispersed in Outer Apra Harbor and on bank shelf locations in Haputo ERA (Figure 3.12d). Like with other corals taxa, no comparison was made to the 2005 NCCOS map (NOAA NCCOS, 2005) because this taxonomic group was not explicitly mapped in that study.

Cover: Seagrass (*Halodule uninervis*)

"Seagrass (*Halodule uninervis*)" (Figure 3.13a) was only present at 1% (3/480) of the training sites (Figure 3.13b). These two sites were located approximately 500 m north of Acapa Point. The Seagrass (*Halodule uninervis*) model showed similar spatial patterns, with the highest likelihood of presence between Dadi Beach and Acapa Point (Figure 3.13c). Probabilities were at or near zero in Haputo ERA. The maximum probability was 55% for H. uninervis. CoV values were lowest (<0.5) in areas with low probabilities (Figure 3.13d). Areas of high CoV (>1) were present on reef features in Haputo ERA, in Inner Apra Harbor, and offshore Dadi Beach. Horizontal artifacts are also visible offshore Dadi Beach in the probability and in the CoV surfaces, which is likely an artifact from the latitude predictor. The predicted patterns in Haputo ERA matched the "Seagrass" class mapped in the 2005 NCCOS map (NOAA NCCOS, 2005). However, these mapped and predicted seagrass distributions differed offshore Dadi Beach. Specifically, the 2005 map showed "Seagrass" nearshore Dadi Beach, whereas the 2023 map predicted "*Halodule uninervis*" mainly offshore Acapa Point.

Cover: Algae (Crustose Coralline)

Crustose coralline algae (Figure 3.14a) was present at 23% (112/480) of the training sites (Figure 3.14b). This taxonomic group was most common in Haputo ERA and from Gab Beach around Point Udall to Acapa Point. The remaining presences were distributed in the middle of Outer Apra Harbor. No presences were documented in Sasa Bay or Inner Apra Harbor. The Algae (Crustose Coralline) model showed similar spatial patterns, with the highest likelihood of presence on bank shelf locations in Haputo ERA and around Orote Peninsula (Figure 3.14d). Probability of occurrence was lowest or zero south of Cabras Island, in Sasa Bay, and in Inner Apra Harbor. The maximum probability was 99% for crustose coralline algae overall. CoV values were lowest (<0.25) in places with high probabilities of occurrence (Figure 3.14e), indicating lower uncertainty for places where it is very likely to be present. The spatial patterns differed from than the distributions of the "Coralline Algae" classes in the 2005 NCCOS map (NOAA NCCOS, 2005). Specifically, the Algae (Crustose Coralline) model predicted these habitats are likely to be present more widely in the project areas than in the 2005 maps.



Live foliose corals inside Apra Harbor. Credit: NOAA NCCOS

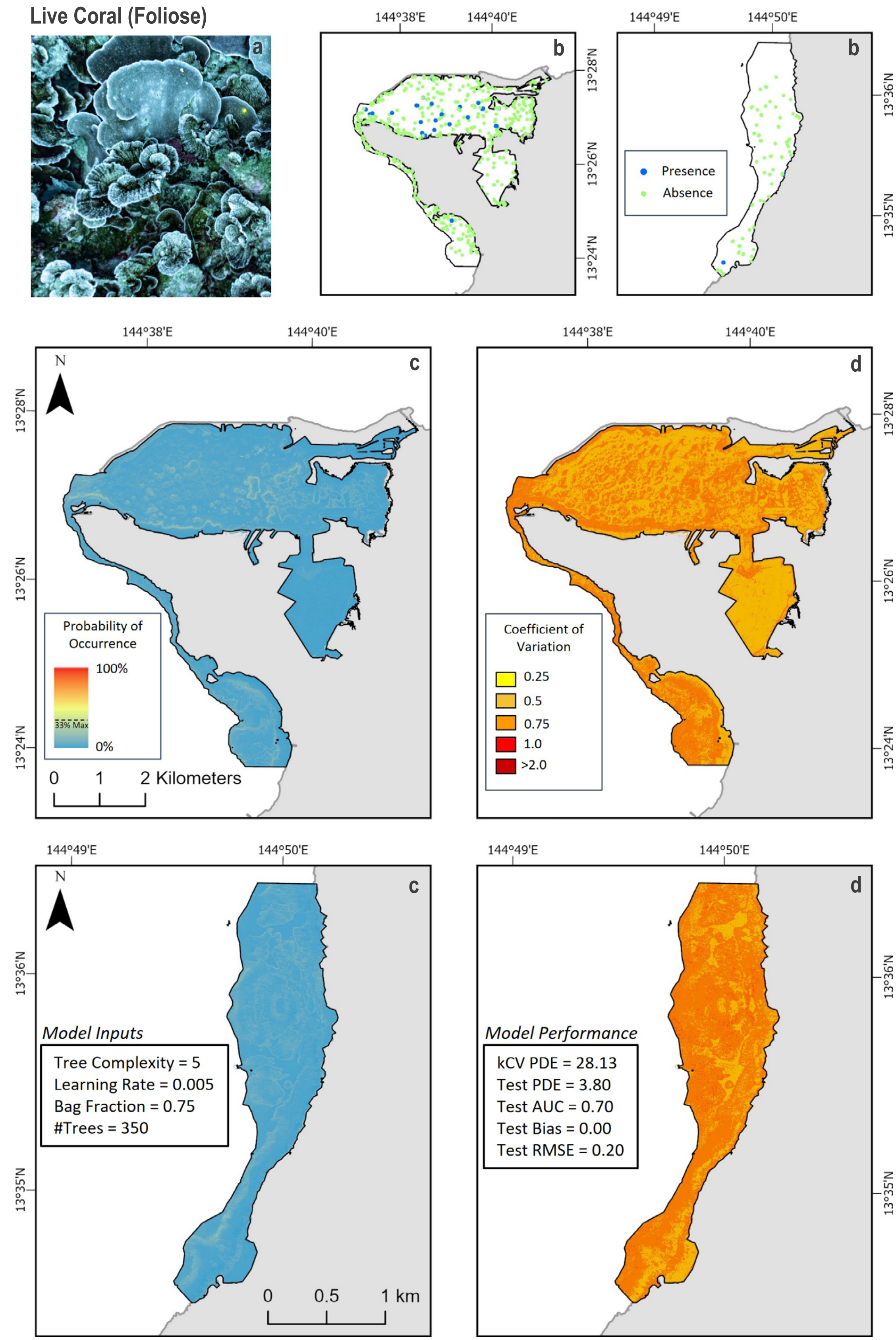


Figure 3.11. Predicted presence of "Live Coral (Foliose)." Figure panels depict: a) a photo of foliose corals; b) maps denoting their presences and absences of foliose corals in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

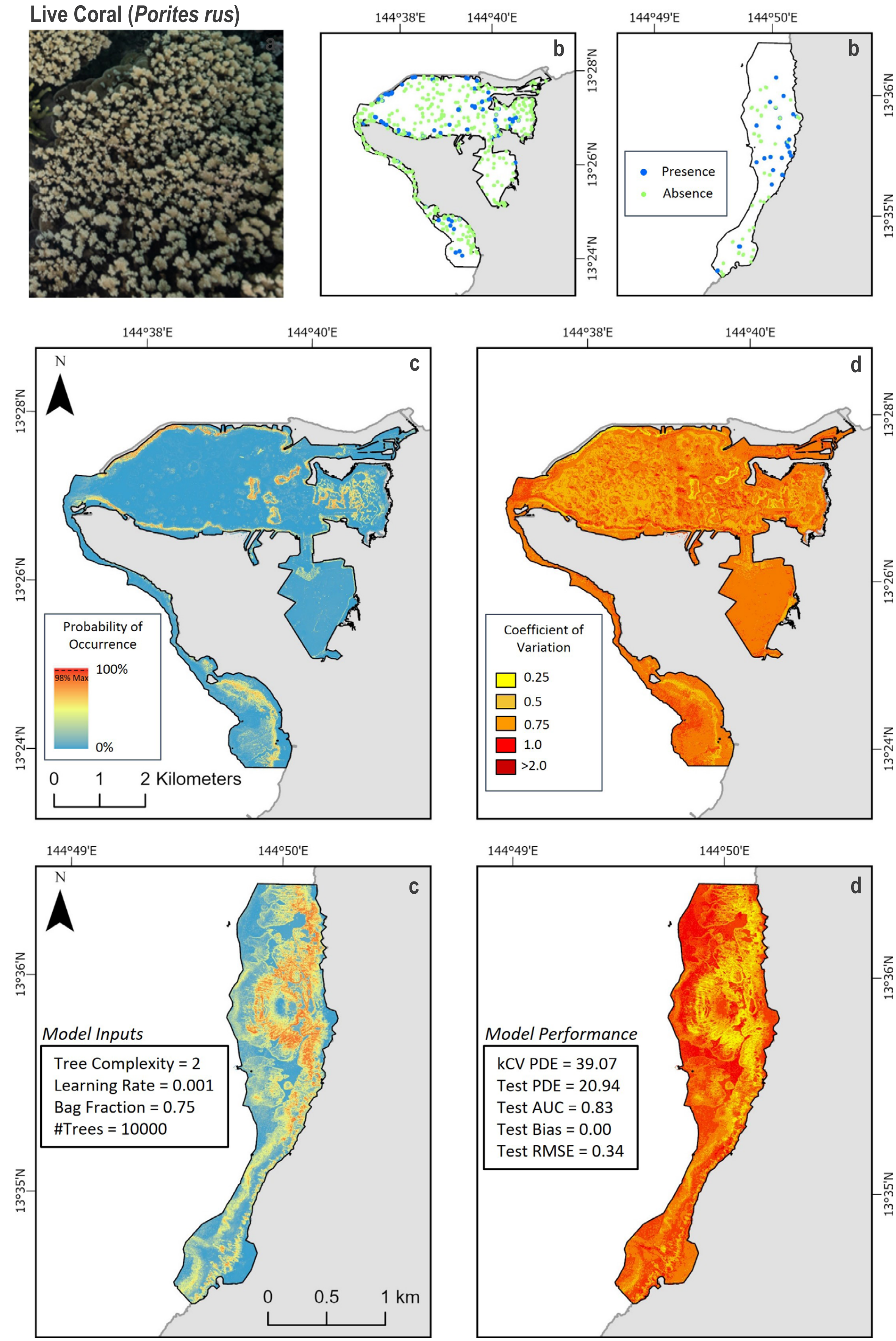


*Figure 3.12. Predicted presence of "Live Coral (*Porites rus*)." Figure panels depict: a) a photo of* Porites rus*; b) maps denoting their presences and absences of* Porites rus *in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.*

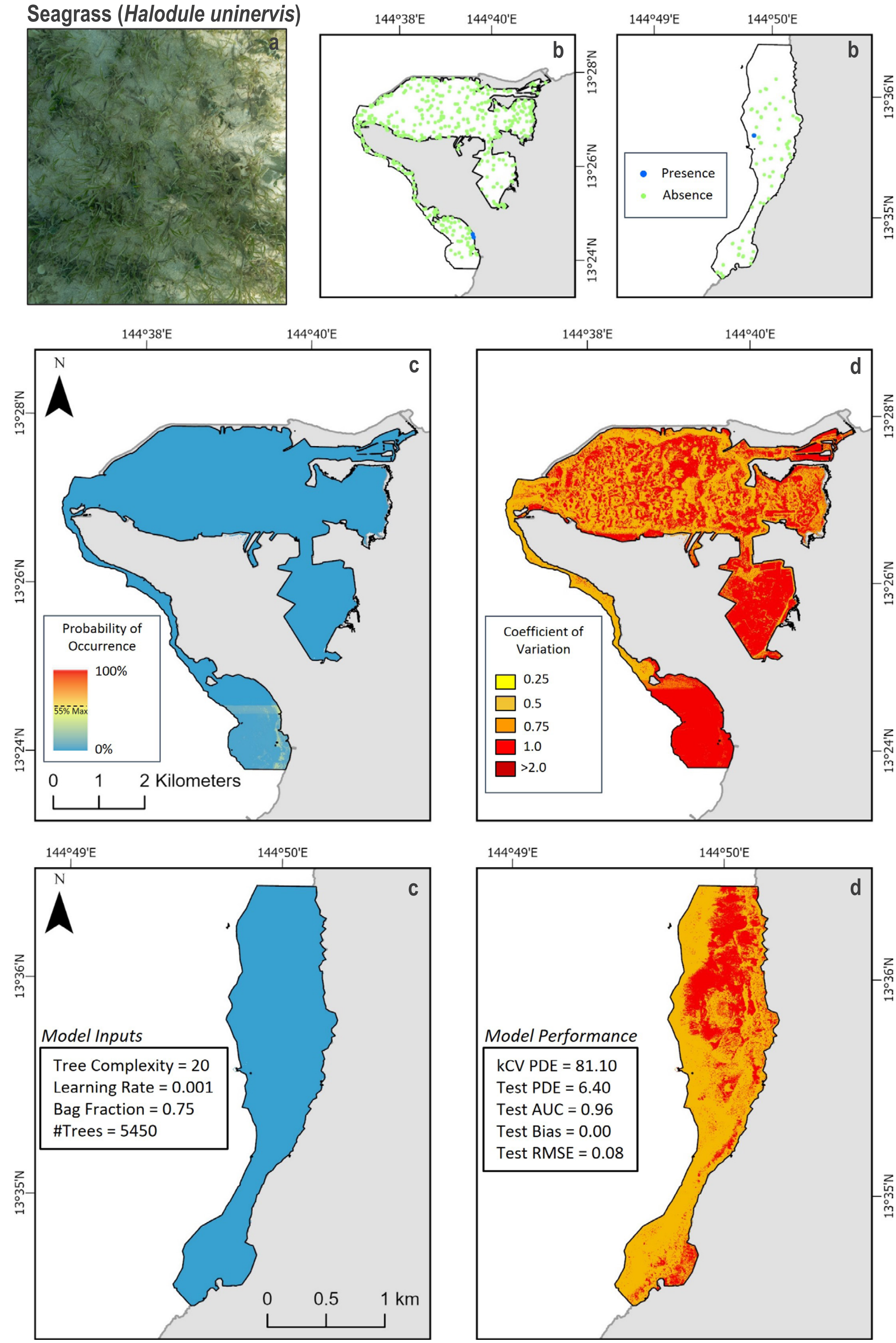


Figure 3.13. Predicted presence of "Seagrass, Halodule uninervis.*" Figure panels depict: a) a photo of* Halodule uninervis*; b) maps denoting the presences and absences of* Halodule uninervis *in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.*

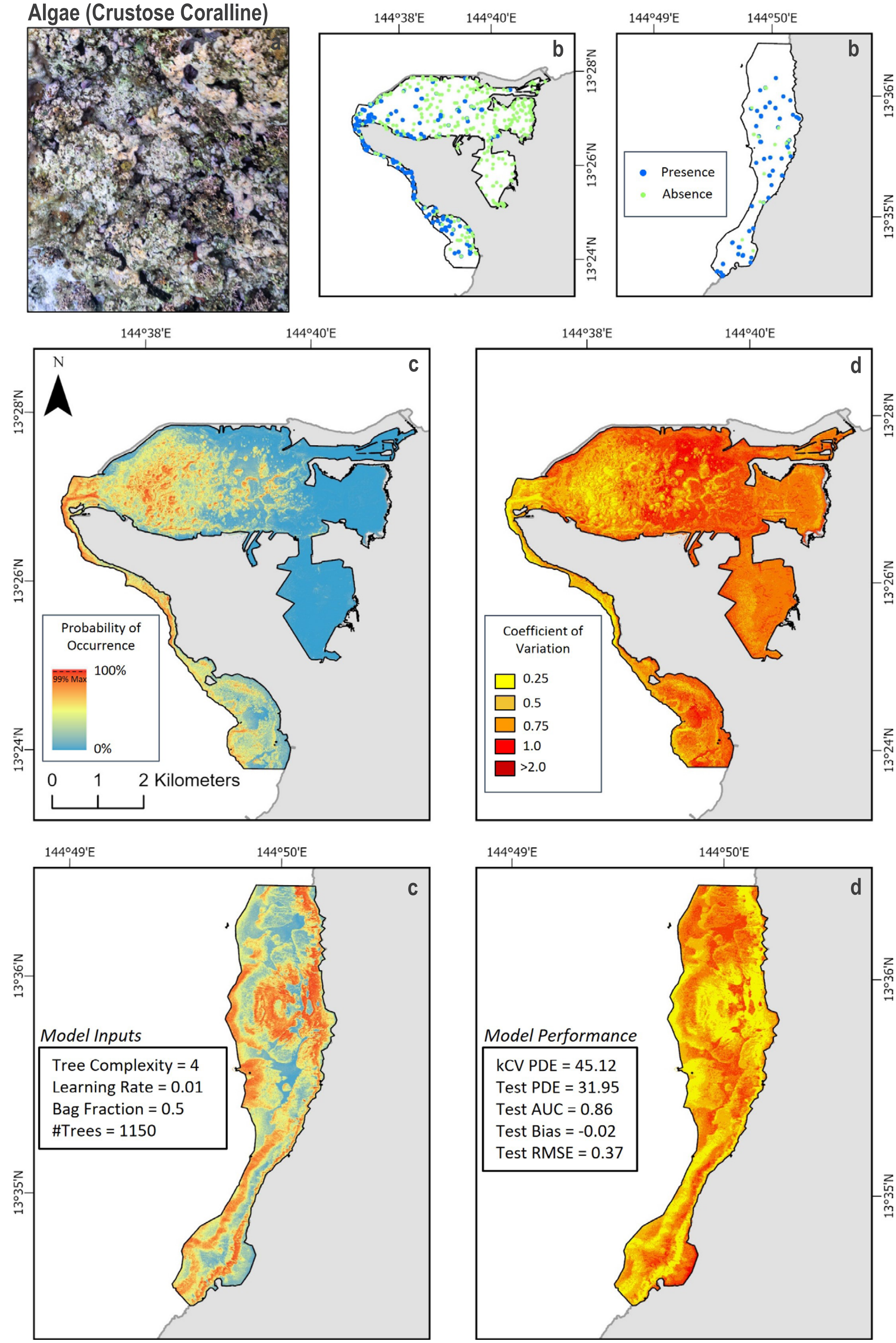


Figure 3.14. Predicted presence of "Algae (Crustose Coralline)" (CCA). Figure panels depict: a) a photo of CCA habitat; b) maps denoting the presences and absences of CCA in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

Cover: Algae (*Halimeda* spp.)

Halimeda algae (Figure 3.15a) was present at 32% (155/480) of the training sites (Figure 3.15b). This taxonomic group was most common in Haputo ERA and from Gab Beach around Point Udall to Acapa Point. The remaining presences were distributed along the Glass Breakwater and in the middle of Outer Apra Harbor. Fewer *Halimeda* presences were documented in Sasa Bay or Inner Apra Harbor. The Algae (*Halimeda* spp.) model showed similar spatial patterns, with the with the highest likelihood of presence in Haputo ERA and from Tipalao Bay to Acapa Point (Figure 3.15c). Probabilities of occurrence were lower in Sasa Bay and in Inner Apra Harbor. The maximum probability was 85% for *Halimeda* overall. CoV values were lowest (<0.25) in places with high probabilities of occurrence (Figure 3.15d), indicating lower uncertainty for places where it is likely to be present. No comparison was made to the 2005 NCCOS map (NOAA NCCOS, 2005) because *Halimeda* was not explicitly mapped in 2005.

Cover: Algae (Turf)

Turf algae (Figure 3.16a) was the most common habitat, and was present at 68% (328/480) of the training sites (Figure 3.16b). This taxonomic group was distributed throughout the project area, with fewer occurrences in Inner Apra Harbor. The Algae (Turf) model showed similar spatial patterns, with the highest likelihood of presence on reefs in Haputo ERA, Outer Apra Harbor, and from Point Udall south to Acapa Point (Figure 3.16c). Probability of occurrence was lowest in places predicted to have majority unconsolidated sediments, including Inner Apra Harbor and non-reef areas in eastern Outer Apra Harbor. The maximum probability of occurrence was 98% for turf algae. CoV values were lowest (<0.25) in places with high probabilities of occurrence (Figure 3.16d) and highest (>0.75) in places with low probabilities of occurrence. This turf algae probability of occurrence prediction partially matched the distributions of the “Turf” biological cover class in the 2005 NCCOS map (NOAA NCCOS, 2005). The notable exceptions are deeper (>25 m) areas in Apra Harbor and from Apuntua Point to Acapa Point, where turf algae was predicted by the 2023 map but not in the 2005 map.

Cover: Algae (Other)

Other types of algae (not listed above) (Figure 3.17a) were present at 63% (303/480) of the training sites (Figure 3.17b). This taxonomic group was common throughout Haputo ERA and throughout Outer Apra Harbor and in Sasa Bay. Fewer presences were documented near the shoreline in Sasa Bay and in Inner Apra Harbor. The Algae (Other) model showed similar spatial patterns, with the with the moderate likelihood of presence in Haputo ERA and in Outer Apra Harbor (Figure 3.17c). Probabilities of occurrence were lowest in places predicted to have soft substrates, including Inner Apra Harbor and non-reef areas in eastern Outer Apra Harbor and Sasa Bay. The maximum probability of occurrence was 83% for other types of algae. CoV values were generally moderate (>0.5) throughout the project areas, with the highest values (>0.75) in Inner Apra Harbor (Figure 3.17d). This probability of occurrence prediction did not match the distributions of the “Macroalgae” biological cover class in the 2005 NCCOS map (NOAA NCCOS, 2005). The notable exceptions where the maps did match are areas around Dry Dock Island and Port Authority Beach.

Cover: Sponge

Sponges (Figure 3.18a) were present at 25% (119/480) of the training sites (Figure 3.18b). This taxonomic group was most common from Point Udall to Acapa Point. Elephant ear sponges (*Ianthella basta*) were also found frequently in Outer Apra Harbor. The remaining presences were distributed throughout Haputo ERA. Few presences were documented in Sasa Bay or inside Inner Apra Harbor. The Sponge model showed similar spatial patterns, with the highest likelihood of presence in Outer Apra Harbor from Dry Dock Island to the Harbor mouth (Figure 3.18c). Probability of occurrence was lowest east of Dry Dock Island south of Cabras Island, Sasa Bay, and Inner Apra Harbor, and in soft sediments in Haputo ERA. The maximum probability was 79% for “Sponge.” CoV values were lowest (<0.25) in places with high probabilities of occurrence (Figure 3.18d), and highest (>0.25) in locations with moderate to low probabilities. No comparison was made to the 2005 NCCOS map (NOAA NCCOS, 2005) because sponges were not explicitly mapped in 2005.

Cover: Bare

Locations without biological cover (i.e., “Bare”) (Figure 3.19a) were common and widely distributed in the project areas. “Bare” cover was present at 71% (341/480) of the training sites (Figure 3.19b). The Bare model showed similar spatial patterns, with the lowest likelihood of this habitat being present on hard bottom in Apra Harbor and Haputo ERA (Figure 3.19c). Probability of occurrence for “Bare” cover was highest in most other soft bottom substrates inside the project areas. The maximum probability was 97% for bare substrate. CoV values were lowest (<0.25) in locations with high probabilities of occurrence (Figure 3.19d) throughout the project areas. The highest (>0.25) CoV values were located along the fore reef in Haputo ERA, and in Apra Harbor from Kilo Wharf around Point Udall to Tipalao Bay. This “Bare” probability of occurrence prediction broadly matched the distributions of the “Uncolonized” class in the 2005 NCCOS map (NOAA NCCOS, 2005). The notable exceptions are from Point Udall to Dadi Beach and in southern Haputo ERA, where the 2005 NCCOS map did not depict any uncolonized substrates.

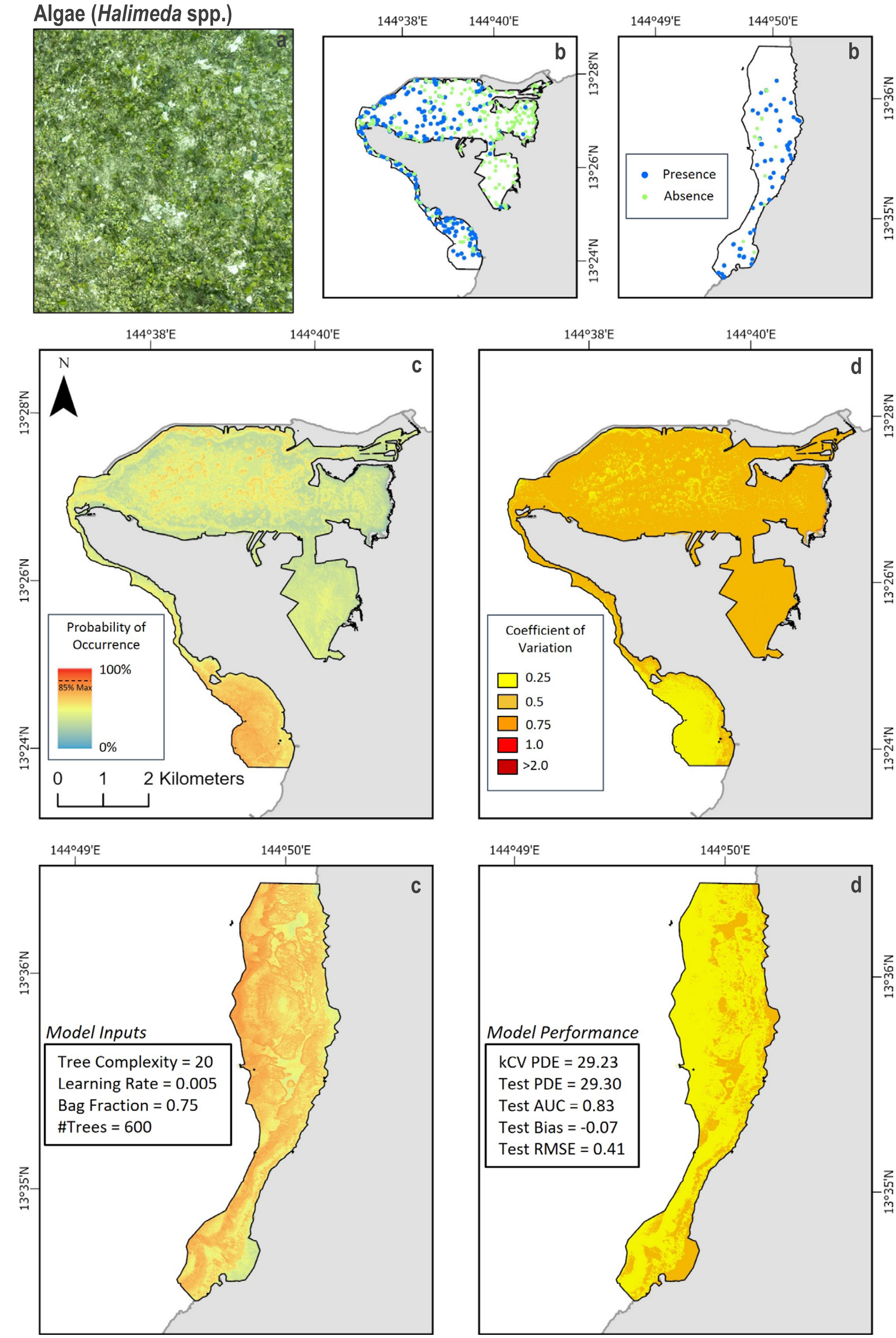


*Figure 3.15. Predicted presence of "Algae (*Halimeda *spp.)." Figure panels depict: a) a photo of* Halimeda *habitat; b) maps denoting the presences and absences of* Halimeda *in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.*

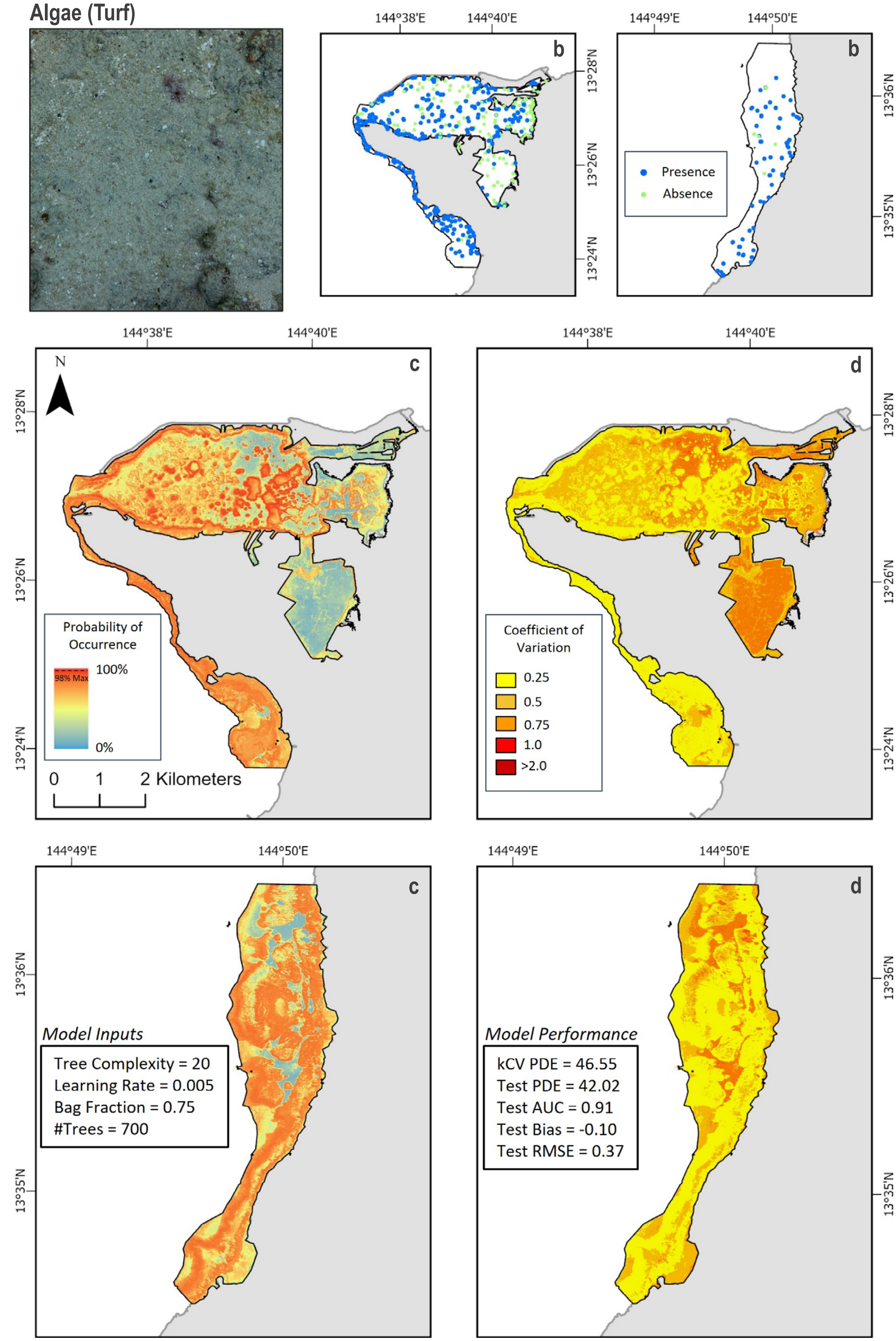


Figure 3.16. Predicted presence of “Algae (Turf).” Figure panels depict: a) a photo of turf algae habitat; b) maps denoting the presences and absences of turf algae in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

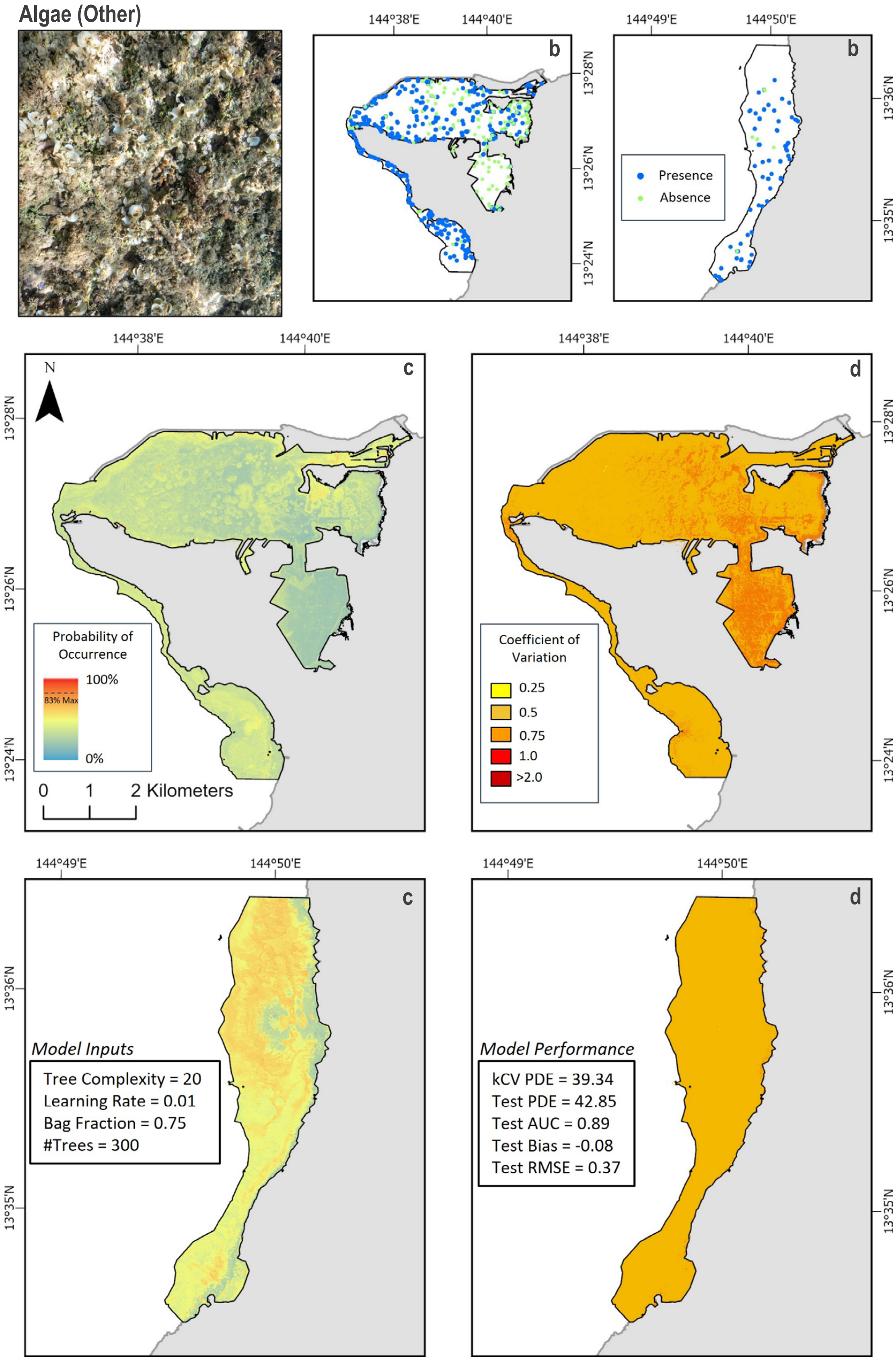


Figure 3.17. Predicted presence of "Algae (Other)." Figure panels depict: a) a photo of other algae habitat; b) maps denoting the presences and absences of other algae in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

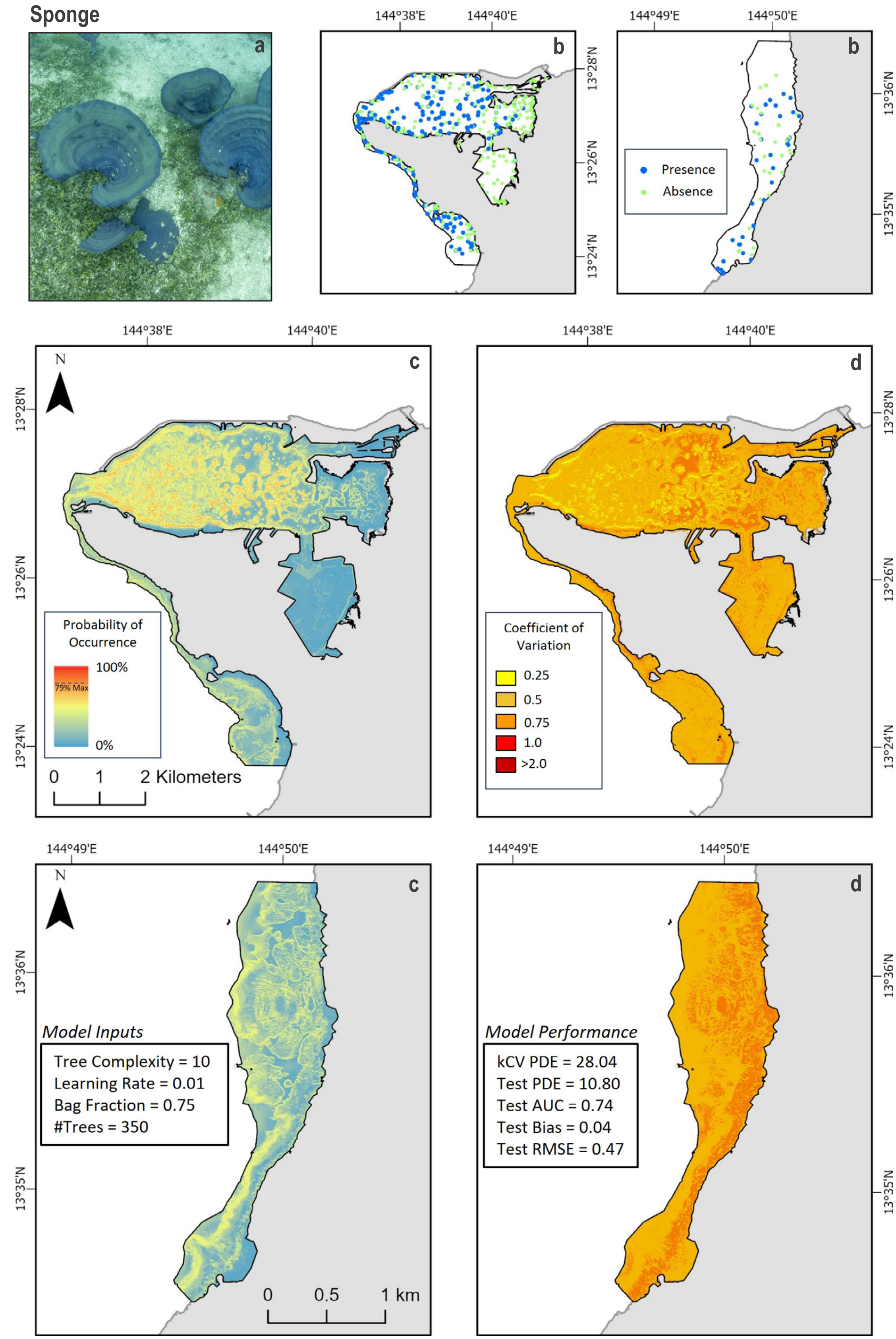


Figure 3.18. Predicted presence of "Sponge." Figure panels depict: a) a photo of sponge habitat; b) maps denoting the presences and absences of sponges in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

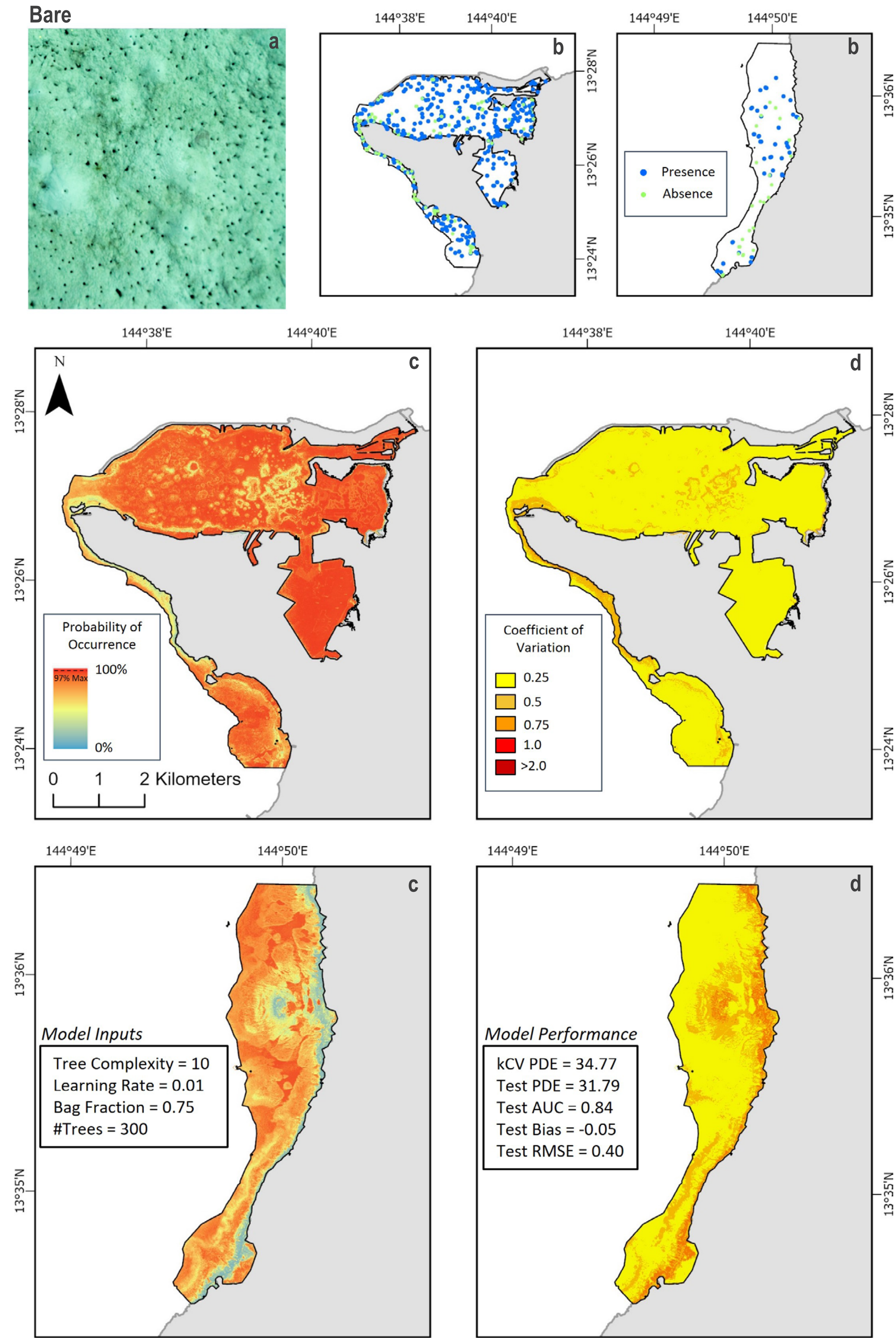


Figure 3.19. Predicted presence of "Bare" cover. Figure panels depict: a) a photo of mixed bare cover habitat; b) maps denoting the presences and absences of bare cover in the training and validation data; c) maps denoting the predicted average probability of occurrence; and d) maps denoting the coefficient of variation. The insets in the bottom panels show the input parameters used to create the model (left), and the performance of the model (right). kCV = k-fold cross validation; PDE = percent deviance explained; AUC = area under the curve; RMSE = root mean square error.

Other Habitats Not Predicted

In addition to the cover types above, other biological organisms were identified and observations made using the underwater photographs. These organisms and observations were not modeled because: (1) their prevalence was too low (<1%) to develop reasonable predictions or (2) their model predictions did not meet the minimum performance thresholds (i.e., AUC >0.7 and PDE >0). These rare or absent organisms specifically included species listed under the ESA (i.e., *Acropora globiceps*, *Isopora palifera*, *Acropora retusa*, *Seriatopora aculeata*), seagrass (Halophila), and nuisance species (angel hair algae [*Chaetomorpha vieillardii*] and crown-of-thorns sea stars [*Acanthaster planci*]). One crown-of-thorns was photographed at a 13-m depth in Haputo ERA (Figure 3.20). No ESA corals or angel hair algae were documented in the project areas. The presences of coral bleaching or paling, crown-of-thorns scarring, and marine debris were also recorded, but their prevalences were also very low (<1%, 0%, and <4%, respectively) throughout the project areas.

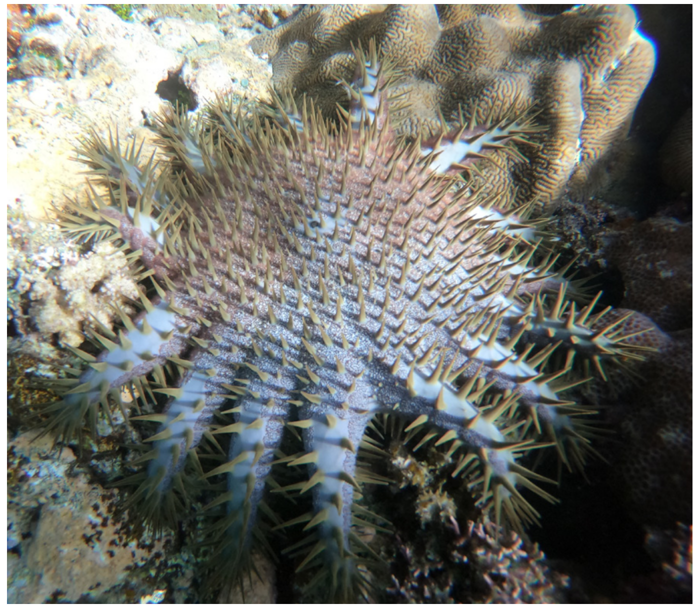


*Figure 3.20. Crown-of-thorns (*Acanthaster planci*) documented in Haputo Ecological Reserve Area in May 2022.*

3.3 Classified Habitat Map

Approximately 21 km^2 of seafloor was characterized around Naval Base Guam and inside Haputo ERA from 0- to approximately 50-m depths. This classified habitat map displays the predicted distribution of seven common combinations of substrate and cover types (Figure 3.21). In Haputo ERA, “Pavement, Mixed Algae” was the most abundant habitat type mapped, comprising 54.5% (1.1 km^2) of the area. The largest continuous patches were located on the fore reef along the coastline. “Sand, Bare” was the next most abundant habitat mapped, comprising 24.8% (0.5 km^2) of the area. Large, continuous patches of bare sand were mainly offshore in the northern area of the reserve. Smaller patches of pavement were also located along the reef tract, along with patches of “Upright Dead Coral Reef, Mixed Algae” (15.1% or 0.3 km^2) and “Live Coral Reef, Live Coral” (5% or 0.1 km^2). These habitat types were the third and fourth most abundant habitats, respectively, mapped in the Haputo ERA. “Sand, Mixed Algae” was the least abundant (0.6% or 0.01 km^2) habitat overall.

In Apra Harbor and from Point Udall to Acapa Point, “Sand, Bare” was the most abundant habitat type mapped, comprising 42.3% (8.2 km^2) of the area. The largest continuous patches were in the eastern portion of Outer Apra Harbor, including Sasa Bay and south of Cabras Island. “Pavement, Mixed Algae” was the next most abundant habitat mapped, comprising 35.9% (6.9 km^2) of the area. Large, continuous patches of pavement covered by algae were concentrated nearshore from San Luis Beach around Point Udall to Acapa Point. Smaller patches of pavement were also located in the western half of Outer Apra Harbor. “Mud, Bare” was the third most abundant habitat (8% or 1.5 km^2), which was concentrated nearshore Sasa Bay and mixed with bare sand in Inner Apra Harbor. “Sand, Mixed Algae” was the fourth most abundant habitat (5.2% or 1 km^2) and was often located where bare mud transitioned to bare sand. “Upright Dead Coral Reef, Mixed Algae” and “Live Coral Reef, Live Coral” were the fifth (4.2% or 0.8 km^2) and sixth (3.0% or 0.6 km^2) most abundant habitats (respectively). These habitats were often co-located on the fore reef along the inside perimeter of Outer Apra Harbor, and from Tipalao Bay to Acapa Point. “Mud, Mangrove” was the least abundant habitat (1.3% or 0.3 km^2), found only in nearshore areas in Sasa Bay and Inner Apra Harbor.

3.3.1 Map Accuracy

The relative prevalence and proportions of the seven habitat types were very similar (±5%) in both the training data and the classified habitat map. Agreement between the training data and classified habitat map suggests the BCTs were able to describe the relationships among the habitats and environmental predictors reasonably well. The notable exceptions were the “Upright Dead Coral Reef, Mixed Algae” and “Sand, Bare” habitats. “Sand, Bare” was more prevalent in the classified habitat map (40.7%) than in the training data (19.4%). Upright dead reef was conversely less prevalent in the classified map (5.2%) than in the training data (13.9%).

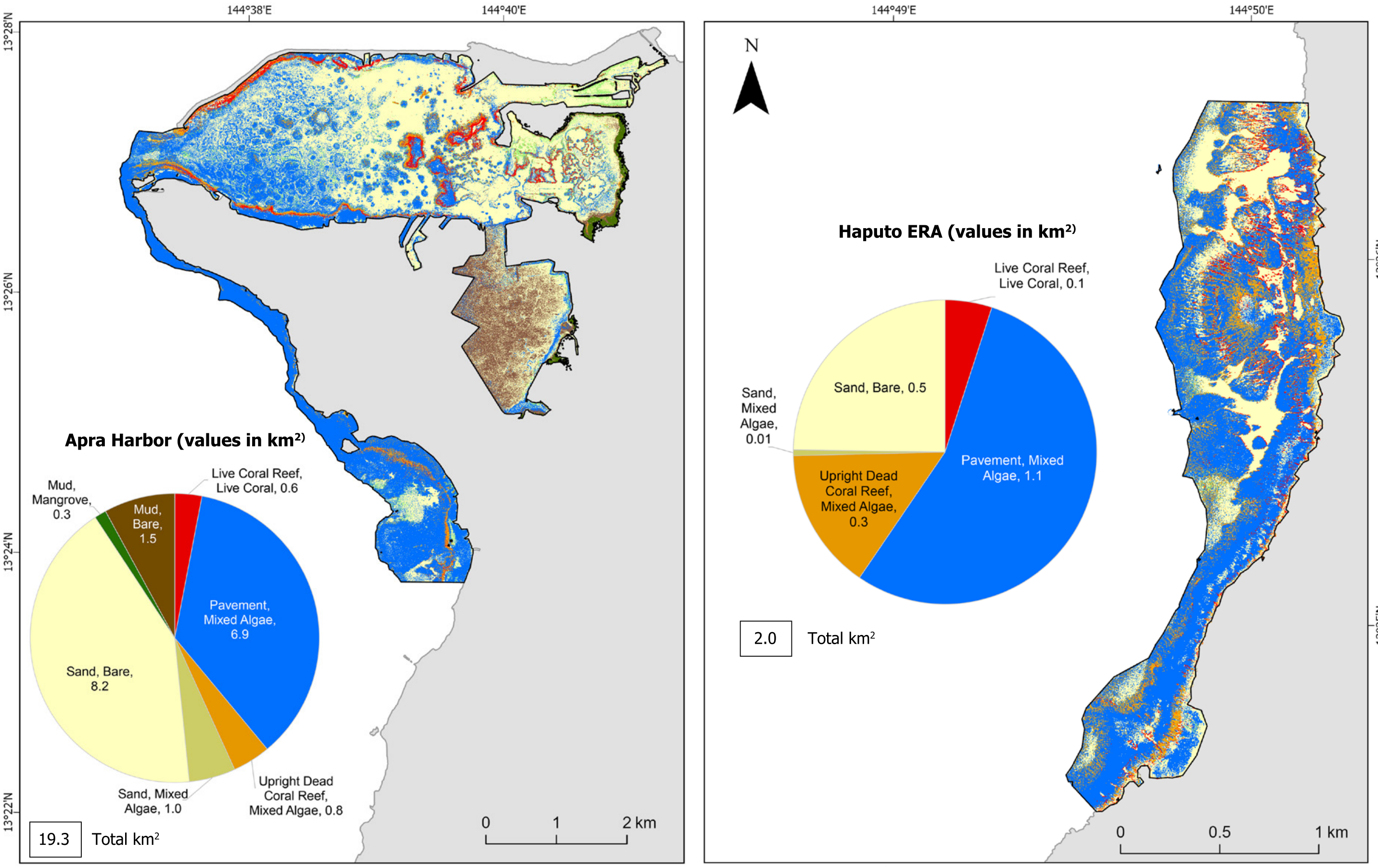


Figure 3.21. This figure depicts the 7 benthic habitats mapped throughout Apra Harbor (left) and Haputo Ecological Reserve Area (ERA; right). The numbers inside the legend denote the amount of area (km2) occupied by each habitat class.

The overall accuracy and tau value for the classified habitat map (quantified using the validation points) was high at 86.6% and 0.84 ±0.04, respectively (Table 3.1). The overall accuracy was very similar after correcting for proportional biases (86.1% ± 4.0% at the 95% confidence level; Table 3.2). The UAs were also high, ranging from 80% to 100% for the individual habitat classes. Most habitat misclassifications were evenly distributed in the confusion table, suggesting that the habitat characterization process did not consistently confuse the majority of substrate and cover type pairs. The notable exceptions were the three hard substrate categories ("Live Coral Reef," "Upright Dead Reef," and "Pavement"), which were confused with each other (n = 8 and 9) more than with the soft bottom habitat categories. Similarly, "Sand, Bare" and "Sand, Mixed Algae" were more often confused with each other (n = 7) than with any other categories. Despite these class-specific biases, the above overall map accuracies are similar to the other benthic habitat maps created by NOAA NCCOS in the Pacific Region (NOAA NCCOS, 2005; Battista et al., 2007; Kendall et al., 2017). As a result, this habitat map can be used with high levels of confidence for a variety of research and management applications.

3.4 Map Applications

3.4.1 Using the Map Products

Spatial and spectral resolution of satellite sensors, computing power, and model-based mapping techniques have advanced considerably in the last decade, and the map products created here take maximum advantage of those improvements to preserve the fine-scale heterogeneity, habitat gradients, and smaller features present in the real landscape.

These new maps were also designed to be flexible, scalable, and customizable to suit specific applications and user needs. Users may apply spatial filters to change the map scale (Kendall and Miller, 2008), enhance dominant or important habitat types, or smooth out variability in heterogeneous areas. Habitat classes can also be aggregated into broader categories (e.g., hard bottom instead of multiple types of reef substrates) or translated into other classification systems (e.g., NOAA NCCOS [2005] or CMECS [2023]) for qualitative comparisons.

Table 3.1. The confusion matrix for the classified habitat map. Observed habitats at validation sites are listed as columns, and corresponding predicted habitats, as rows. Cell values are the number of matches (along the gray diagonal) or mismatches (off diagonal) between the two. nj = row total; ni = column total; OA = overall accuracy; CI = confidence interval; UA = user's accuracy; PA = producer's accuracy.

Predicted (j) \ Observed (i)	Live Coral Reef, Live Coral	Pavement, Mixed Algae	Sand, Mixed Algae	Upright Dead Coral Reef, Mixed Algae	Sand, Bare	Mud, Bare	Mud, Mangrove	n_j	User's Accuracy (%)
Live Coral Reef, Live Coral	**16**	0	0	0	1	0	0	**17**	94%
Pavement, Mixed Algae	9	**106**	6	8	2	1	0	**132**	80%
Sand, Mixed Algae	0	0	**17**	1	1	0	0	**19**	89%
Upright Dead Coral Reef, Mixed Algae	1	0	2	**32**	0	0	0	**35**	91%
Sand, Bare	3	2	7	0	**80**	0	0	**92**	87%
Mud, Bare	0	0	0	0	0	**20**	0	**20**	100%
Mud, Mangrove	0	0	0	0	0	0	**13**	**13**	100%
n_i	29	108	32	41	84	21	13	**328**	
PA (%)	55%	98%	53%	78%	95%	95%	100%	OA = 86.6% Tau = 0.84 CI (±) = 0.04	

Table 3.2. The confusion matrix for the classified habitat map corrected for proportional biases. Observed habitats at validation sites are listed as columns, and corresponding predicted habitats, as rows. Cell values are the number of matches (along the gray diagonal) or mismatches (off diagonal) corrected for proportional bias; OA = overall accuracy; CI = confidence interval; UA = user's accuracy; PA = producer's accuracy; πj = observed habitat proportion, pi = predicted habitat proportion.

Predicted (j) \ Observed (i)	Live Coral Reef, Live Coral	Pavement, Mixed Algae	Sand, Mixed Algae	Upright Dead Coral Reef, Mixed Algae	Sand, Bare	Mud, Bare	Mud, Mangrove	π_j	UA	UA CI (±)
Live Coral Reef, Live Coral	**0.030**	0.000	0.000	0.000	0.002	0.000	0.000	0.032	94.1%	0.86%
Pavement, Mixed Algae	0.026	**0.302**	0.017	0.023	0.006	0.003	0.000	0.377	80.3%	3.33%
Sand, Mixed Algae	0.000	0.000	**0.043**	0.003	0.003	0.000	0.000	0.048	89.5%	1.43%
Upright Dead Coral Reef, Mixed Algae	0.001	0.000	0.003	**0.048**	0.000	0.000	0.000	0.052	91.4%	1.04%
Sand, Bare	0.013	0.009	0.031	0.000	**0.354**	0.000	0.000	0.407	87.0%	3.58%
Mud, Bare	0.000	0.000	0.000	0.000	0.000	**0.072**	0.000	0.072	100.0%	0.00%
Mud, Mangrove	0.000	0.000	0.000	0.000	0.000	0.000	**0.012**	0.012	100.0%	0.00%
ρ_i	0.071	0.311	0.094	0.073	0.364	0.075	0.012			
PA (%)	42.7%	97.2%	45.5%	65.4%	97.2%	96.2%	100.0%	OA = 86.1% CI (±) = 4.0%		
PA CI (±)	14.14%	9.01%	13.89%	15.47%	8.10%	7.29%	0.00%			

Predictions for individual substrate and cover types may also be converted from the continuous probability values ideal for examining gradients into classified categories to better characterize subtle or more dramatic shifts in presence. Users can explore these shifts by customizing and applying their own probability thresholds.

In addition to probability thresholds, users can also explore the impacts of map uncertainty on potential management scenarios and on their decisions (Costa et al., 2019). Here, uncertainty was quantified as CoV. Smaller CoVs indicate that the prediction has higher precision and less uncertainty (and vice versa). CoV can be multiplied by mean probability of occurrence to derive the standard deviation and thereby quantiles and confidence intervals associated with the estimated probabilities in a pixel. For example, if the mean probability is 0.5 and the CoV is 0.1 in a pixel, then $0.5 \times 0.1 = \pm 0.05$ or ± 1 standard deviation (SD). Assuming normally distributed errors, 68.3% of the data will fall between ± 0.05 (or 1 SD), 95.5% of the data will fall between ± 0.1 (or 2 SD), and 99.7% of the data will fall between ± 0.15 (or 3 SD).

Testing and changing maps based on their CoV may lead users to different conclusions and courses of action. Defining acceptable levels of uncertainty upfront is critical for users to ensure that they will meet their marine resource goals. It can also help users more confidently identify priority sites, adequately protect habitats, convey the range of potential outcomes, and ensure that limited resources are used as efficiently as possible (Margules and Pressey, 2000; Nicholson and Possingham, 2007; Tulloch et al., 2013). That said, it is important to note that high CoVs may occur in areas of predicted absence where mean probabilities are extremely small (values are close to 0). Users should be aware that high CoVs in those locations do not necessarily indicate high uncertainty, and these areas should be reviewed alongside the mean predictions to avoid misinterpretation.

3.4.2 Informing Management Decisions

Submerged lands in and around Apra Harbor are used by the Navy for a variety of training exercises and activities, which have the potential to impact coral reef ecosystems. Naval activities and actions that potentially affect coral reef ecosystems must be mitigated under Executive Order 13089. These map products will be used by NAVFAC Marianas to comply with this executive order and to guide how best to minimize impacts to important habitats in Apra Harbor and Haputo ERA.

In addition to supporting NAVFAC Marianas, the map products described here were designed with these and other potential management uses in mind. Notably, these products may inform other local marine monitoring and management decisions, such as identifying and monitoring nuisance species (e.g., *Chaetomorpha vieillardii* and *Acanthaster planci*) (Guam BSP, 2018), quantifying the economic value of coral reefs (van Beukering, 2007), calculating damage and costs following ship grounding or other impacts (Brown, 2015), monitoring habitat changes through time (Pendleton et al., 2005), minimizing development impacts to important habitats (Nelson et al., 2016), designing sampling plans for monitoring or scientific studies (Guam and NOAA CRCP, 2010), and conducting education and outreach.

Regardless of the application, the best way to access and use these highly resolved maps is through GIS or other software that allows users to zoom in and out as needed. The GIS-ready products from this project are listed below:

1. Map of classified benthic habitats in and around Apra Harbor and in Haputo ERA from 0- to 50-m depths;
2. Maps of the predicted occurrence of 19 substrate and biological cover types;
3. Orthorectified, atmospheric- and water column–corrected satellite images;
4. Maps depicting the depth, roughness, hardness, and topography of the seafloor. These data for Inner Apra Harbor are restricted and are not publicly available.
5. Underwater photographs and annotations used to train model development and validate their performance and accuracy;
6. A technical report (this document) describing the methods, results, and limitations for scientific and management applications of these products.

These GIS products are freely available for download here:
https://coastalscience.noaa.gov/project/characterizing-submerged-lands-around-navy-base-guam-cnmi/
https://coastalscience.noaa.gov/project/characterizing-benthic-habitats-in-haputo-ecological-reserve-area-guam/

If users do not have GIS software, tiled versions of the satellite imagery (Appendix C) and classified habitat maps (Appendix D) are available for printing, and a data viewer is available online for viewing and querying these habitat products without any specialized software: https://experience.arcgis.com/experience/7b6c0e7164234182985a89d5b5703475

References

Agisoft. (2023). Agisoft Metashape Professional Edition. https://www.agisoft.com/features/professional-edition/

Allen, S. D., and Bartram, P. K. (2008). Guam as a fishing community. https://repository.library.noaa.gov/view/noaa/3583

Amesbury, S., Bonito, V., Chang, R., Kirkendale, L., Meyer, C., Paulay, G., Ritson-Williams, R., and Rongo, T. (2001). Marine biodiversity resource survey and baseline reef monitoring survey of the Haputo Ecological Reserve Area, COMNAVMARIANAS, Guam. Report and Interactive GIS Document prepared for U.S. Dept. of Defense, COMNAVMARIANAS. 111 pp. and CD-ROM.

Battista, T. A., Costa, B. M., and Anderson, S. M. (2007). Shallow-water benthic habitats of the Republic of Palau. NOAA Technical Memorandum NOS NCCOS 59.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. I. (1984). Classification and regression trees. Taylor & Francis Group. https://doi.org/10.1201/9781315139470

Brown, V. (2015). Assessment of Daiki Maru Grounding, Spanish Steps, Orote Peninsula, Outer Apra Harbor, Guam. https://www.coris.noaa.gov/metadata/records/html/194_FY14_Metadata_DaikiMaru_2014.html

Burdick, D. (2005). Guam coastal atlas. https://www.uog.edu/_resources/files/ml/technical_reports/114Burdick_2005_UOGMLTechReport114.pdf

Burdick, D., Brown, V., Asher, J., Gawel, M., Goldman, L., Hall, A., Kenyon, J., Leberer, T., Lundblad, E., McIlwain, J., Miller, J., Minton, D., Nadon, M., Pioppi, N., Raymundo, L., Richards, B., Schroeder, R., Schupp, P., Smith, E., and Zgliczynski, B. (2008). The state of coral reef ecosystems of Guam. In: J. E. Waddell and A. M. Clarke (Eds.), *The state of coral reef ecosystems of the United States and Pacific freely associated states: 2008* (pp. 465–509). NOAA Technical Memorandum NOS NCCOS 73. https://repository.library.noaa.gov/view/noaa/17794

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, 361–445. U.S. Government Printing Office. https://doi.org/10.3133/pp280K

CMECS. (2023). Coastal and Marine Ecological Classification Standard Catalog of Units. https://www.cmecscatalog.org/

Congalton, R. G., and Green, K. (1999). Assessing the accuracy of remotely sensed data: Principles and practices. CRC/Lewis Press. https://doi.org/10.1201/9781420055139

Costa, B., Kendall, M., and McKagan, S. (2018). Managers, modelers, and measuring the impact of species distribution model uncertainty on marine zoning decisions. *PLOS ONE, 13*(10), e0204569. https://doi.org/10.1371/journal.pone.0204569

De'ath, G., and Fabricius, K. E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology, 81*, 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology, 88*, 243–251. https://doi.org/10.1890/0012-9658(2007)88[243:btfema]2.0.co;2

Elith J., Graham, C. H., Anderson, R. P, Dudı´k, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T, Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S. E., Wisz, M. S., and Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography, 29*, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*, 802–81. https://doi.org/10.1111/j.1365-2656.2008.01390.x

ESRI. (2023). ArcIGS Pro (Version 3.x). Environmental Systems Research Institute. Retrieved from http://www.esri.com

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis, 38*, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Green, E. P., Mumby, P. J., Edwards, A. J., and Clark, C. D. (2000). Water column correction techniques. In A. J. Edwards (Ed.), Remote sensing handbook for tropical coastal management, (pp. 121–128). *Coastal Management Sourcebooks, 3*(4). UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000119752

Guam BSP. (2018). November: Reef Invaders. https://bsp.guam.gov/wp-bspcontent/uploads/2018/11/November-GYOR-newsletter_FINAL.pdf

References

HDR and CSA. (2017). R2 Sonic Benthic Habitat Mapping, Outer Apra Harbor, Naval Base Guam. Prepared by: HDR, Hagåtña, Guam, and CSA Ocean Sciences, Inc., Stuart, Florida.

HDR and CSA. (2018). Edgetech 6205 Shallow Water Hydrographic Survey Brief, Apra Harbor, Guam. Prepared by: CSA Ocean Sciences, Inc., Stuart, Florida.

Hijmans, R. J., van Etten, J., Sumner, M., Cheng, J., Baston, D., Bevan, A., Bivand, R., Busetto, L., Canty, M., Fasoli, B., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J. A., Hiemstra, P., Hingee, K., Ilich, A., Karney, C., Mattiuzzi, M., Mosherm, S., Naimi, B., Nowosad, J., Pebesma, E., Lamigueiro, O. P., Racine, E. B., Rowlingson, B., Shortridge, A., Venables, B., and Wueest, R. (2023a). R package, Raster: Geographic data analysis and modeling. https://cran.r-project.org/web/packages/raster/index.html

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J. (2023b). R package, dismo: Species distribution modeling. http://CRAN.R-project.org/package=dismo

Hosmer, D. W., and Lemeshow, S. (2000). Applied logistic regression (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/0471722146

Kendall, M. S., Costa, B., McKagan, S., Johnston, L., and Okano, D. (2017). Benthic habitat maps of Saipan Lagoon. NOAA Technical Memorandum NOS NCCOS 229. https://doi.org/10.7289/V5/TM-NOS-NCCOS-229

Kendall, M. S., and Miller, T. (2008). The influence of thematic and spatial resolution on maps of a coral reef ecosystem. *Marine Geodesy, 31*, 75–102. https://doi.org/10.1080/01490410802053617

Kuhn, M. (2023). R package, Caret: Classification and Regression Training. https://cran.r-project.org/web/packages/caret/

Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics, 17,* 379–383. https://opg.optica.org/ao/abstract.cfm?URI=ao-17-3-379

Ma, Z., and Redmond, R. L. (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing, 61*, 435-439. https://www.asprs.org/wp-content/uploads/pers/1995journal/apr/1995_apr_435-439.pdf

Margules, C. R., and Pressey, R. L. (2000). Systematic conservation planning. *Nature, 405*(6783), 243–253. https://doi.org/10.1038/35012251

Marler, T. E., and Moore, A. (2011). Military threats to terrestrial resources not restricted to wartime: a case study from Guam. *Journal of Environmental Science and Engineering, 5*(9), 1198–1214. https://www.researchgate.net/publication/258227258_Military_threats_to_terrestrial_resources_not_restricted_to_wartime_a_case_study_from_Guam_J

Mumby, P. J., Clark, C. D., Green, E. P., and Edwards, A. J. (1998). Benefits of water column correction and contextual editing for mapping coral reefs. *International Journal of Remote Sensing, 19,* 203–210. https://doi.org/10.1080/014311698216521

Nelson, D. S., McManus, J., Richmond, R. H., King Jr, D. B., Gailani, J. Z., Lackey, T. C., and Bryant, D. (2016). Predicting dredging-associated effects to coral reefs in Apra Harbor, Guam – Part 2: Potential coral effects. *Journal of Environmental Management, 168,* 111–122. https://doi.org/10.1016/j.jenvman.2015.10.025

Nicholson E., and Possingham, H. P. (2007). Making conservation decisions under uncertainty for the persistence of multiple species. *Ecological Applications, 17,* 251–265. https://www.jstor.org/stable/40061991

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. https://repository.library.noaa.gov/view/noaa/420/noaa_420_DS1.pdf

NOAA OCS. (2001). Hydrographic Descriptive Report (Report for W00006, Micronesia-Marianas Islands, Apra Inner Harbor). https://www.ngdc.noaa.gov/nos/W00001-W02000/W00006.html

NOAA COOPS. (2022). Center for Operational Oceanographic Products and Services, Apra Harbor, Guam, United States of America - Station ID: 1630000 [Data set]. https://tidesandcurrents.noaa.gov/stationhome.html?id=1630000

NOAA CRCP. (2018). Coral reef condition: A status report for Guam. https://www.coris.noaa.gov/monitoring/status_report/docs/Guam_status_report_forweb.pdf

NOAA NCCOS. (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. https://cdn.coastalscience.noaa.gov/datasets/e99/us_pac_terr_atlas.pdf

NOAA NGS. (2020). USGS/NOAA topobathy lidar DEM: Guam [Data set]. https://coast.noaa.gov/dataviewer/#/lidar/search/-24001685.50371505,1475081.7718473314,-23906903.58864143,1548461.3190011005/details/9323

References

Paulay, G., Kirkendale, L., Meyer, C., Houk, P., Rongo, T., and Chang, R. (2001). Marine biodiversity resource survey and baseline reef monitoring survey of the Southern Orote Peninsula and North Agat Bay Area, COMNAVMARIANAS. Report and Interactive GIS Document Prepared for U.S. Dept. of Defense, COMNAVMARIANAS. 111 pp. and CD-ROM.

Paulay, G. (2003). Marine Biodiversity of Guam and the Marianas: overview. *Micronesica, 35*(36), 3–25. https://www.researchgate.net/publication/228500677_The_marine_biodiversity_of_Guam_and_the_Marianas

Pendleton, E. A., Thieler, E. R., and Williams, S. J. (2005). *Coastal Vulnerability Assessment of War in the Pacific National Historical Park (WAPA) to Sea-Level Rise.* U.S. Geological Survey. https://doi.org/10.3133/ofr20051056

PIBHMC. (2010). Preliminary hard and soft bottom seafloor substrate map derived from gridded sidescan and bathymetry derivatives at Apra Harbor, Guam US Territory [Data set]. https://www.coris.noaa.gov/metadata/records/html/apra_5m_hardsoft.html

Pierce, J., and Winians, R. (2023). Github: Drop-e. https://github.com/Jordan-Pierce/drop-e

Raymundo, L. J., Burdick, D., Hoot, W. C., Miller, R. M., Brown, V., Reynolds, T., Gault, J., Idechong, J., Fifer, J., and Williams, A. (2019). Successive bleaching events cause mass coral mortality in Guam, Micronesia. *Coral Reefs, 38,* 677–700. https://doi.org/10.1007/s00338-019-01836-2

R Core Team. (2022). R: A language and environment for statistical computing (Version 4.3.x). R Foundation for Statistical Computing. https://www.r-project.org/

Smith, B. D., Donaldson, T. J., Schils, T., Dugger, K., den Haan, J., Myers, R. L., Chau, L., Reyes, A., Chop, K., and Christy, M. (2009). Marine Monitoring at Orote and Haputo Ecological Reserve Areas. Prepared by: SWCA Environmental Consultants, Pasadena, California. https://www.researchgate.net/publication/264824567_Marine_monitoring_at_Orote_and_Haputo_ecological_research_areas

Story, M., and Congalton, R. (1986). Accuracy assessment: A user's perspective. Photogrammetric Engineering and Remote Sensing, 52, 397–399. https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf

Tulloch, V. J., Possingham, H. P., Jupiter, S. D., Roelfsema, C., Tulloch, A. I., and Klein, C. J. (2013). Incorporating uncertainty associated with habitat data in marine reserve design. *Biological Conservation, 162,* 41–51. https://doi.org/10.1016/j.ocecoaman.2013.04.005

U.S. Navy. (2010). Final Environmental Impact Statement, Guam and CNMI Military Relocation, Relocating Marines from Okinawa, Visiting Aircraft Carrier Berthing, and Army Air and Missile Defense Task Force. Joint Guam Program Office, United States Department of the Navy.

U.S. Navy. (2012). Joint Region Marianas Integrated Natural Resources Management Plan: Guam, Tinian, Farallon de Medinilla. Prepared by SWCA Environmental Consultants, Hagåtña, Guam. https://sablan.house.gov/sites/evo-subsites/sablan-evo.house.gov/files/documents/PubRev091912-1_JRM%20INRMP%20091912.pdf

U.S. Navy. (2022). Joint Region Marianas Integrated Natural Resources Management Plan: Guam, Tinian, and Farallon de Medinilla. Prepared by: Cardno GS, Inc., Honolulu, HI.

van Beukering, P., Haider, W., Longland, M., Cesar, H., Sablan, J., Shjegstad, S., Beardmore, B., Liu, Y., and Garces, G. O. (2007). The economic value of Guam's coral reefs. University of Guam Marine Laboratory Technical Report 116. https://www.ncei.noaa.gov/data/oceans/coris/library/NOAA/CRCP/project/1029/econ_value_guam_coral_reefs.pdf

Veron, J. E., DeVantier, L. M., Turak, E., Green, A. L., Kininmonth, S., Stafford-Smith, M., and Peterson, N. (2011). The coral triangle. In *Coral reefs: an ecosystem in transition* (pp. 47–55). https://doi.org/10.1007/978-94-007-0114-4_5

Wusinich-Mendez, Dana. (2010). Guam's coral reef management priorities. https://repository.library.noaa.gov/view/noaa/621

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 (BCT) 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 stage-wise (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 (BRT) 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 stage-wise (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 resampling technique for estimating the statistical uncertainty (precision) in model predictions.

Coefficient of variation (CoV) – 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.

Learning rate (lr) – In a boosting context, the degree to which each base learner 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%. Higher values indicate better model performance. kCV PDE is calculated using the training data and k-fold cross validation. Test PDE is calculated using the validation data only.

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 an 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. Test AUC is calculated using the validation data only.

Resampling – A method of using randomly drawn subsets of data to estimate statistical precision (e.g., variation in model predictions), perform a significance test (e.g., permutation test of predictor importance), or perform model validation (e.g., cross-validation). The term "resampling" can also be used in a geographic information system (GIS) to describe the interpolation methods used to change the resolution of a raster dataset.

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 field data). Test RMSE is calculated using the validation data only.

Spatial autocorrelation – A measure of similarity (correlation) between nearby observations.

Spatial predictive modeling – A modeling technique whereby relationships between environmental predictors and a response variable are estimated for locations with survey data (e.g., underwater photographs). These relationships are then used to predict the response in locations without survey data.

Sensitivity – Also known as the true positive rate, a measure of model performance for binary classification models (e.g., presence versus absence) that measures the proportion of positives that are correctly identified as positives.

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

Validation data – Data that are excluded during model fitting and later used to independently validate the predictive performance of the fitted model and/or accuracy of the classified map.

Training data – Data to which a model is fitted (using kCV) in order to test and optimize 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. In general, greater tree complexity results in fewer iterations needed for model convergence.

Appendix A

Apra Harbor

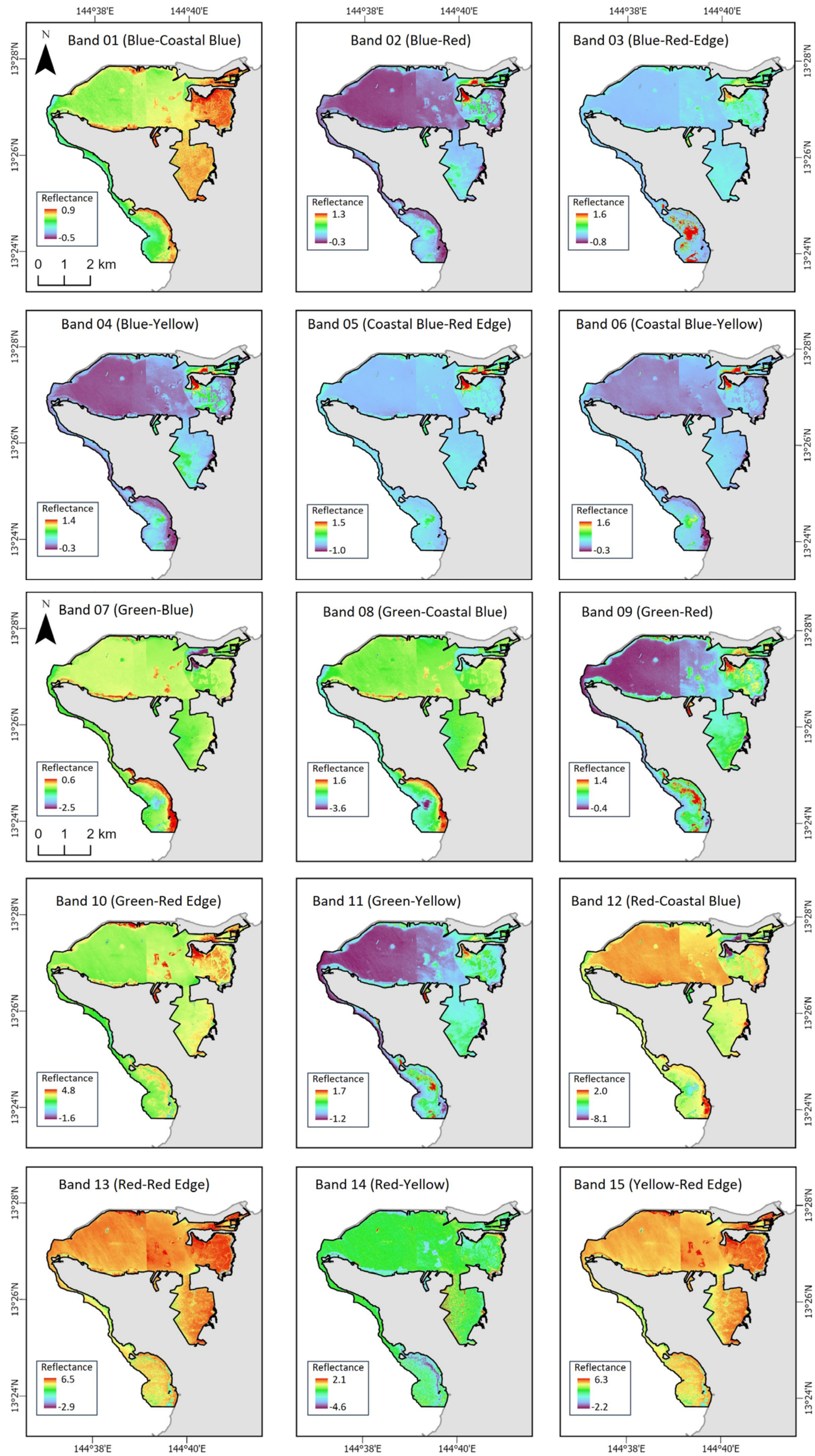


Figure A1. Maps depicting the orthorectified, atmospheric- and water column–corrected WV2 and WV3 band pairs used to create the habitat predictions for Apra Harbor.

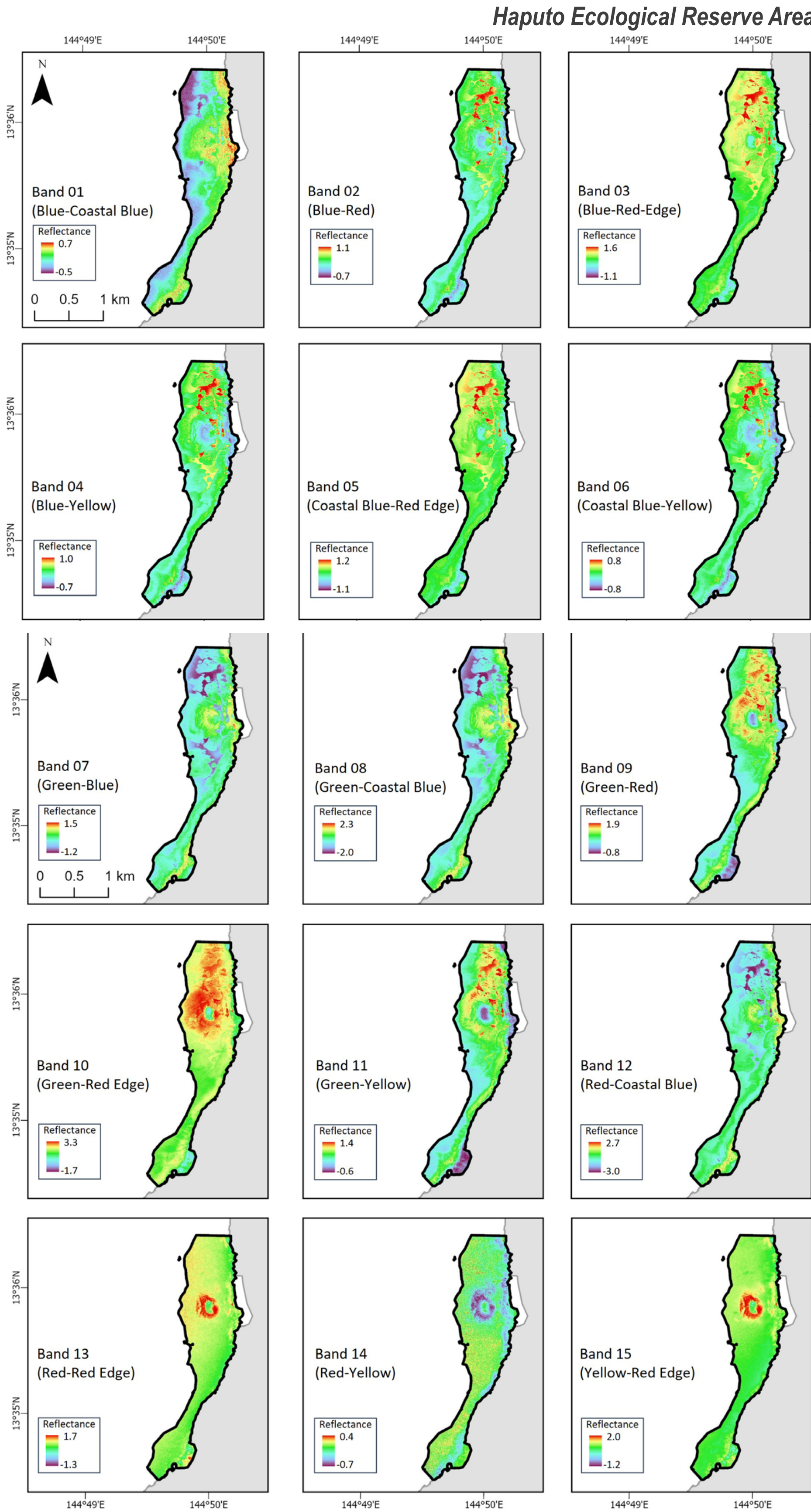


Figure A2. Maps depicting the orthorectified, atmospheric- and water column–corrected WV2 and WV3 band pairs used to create the habitat predictions for Haputo ERA.

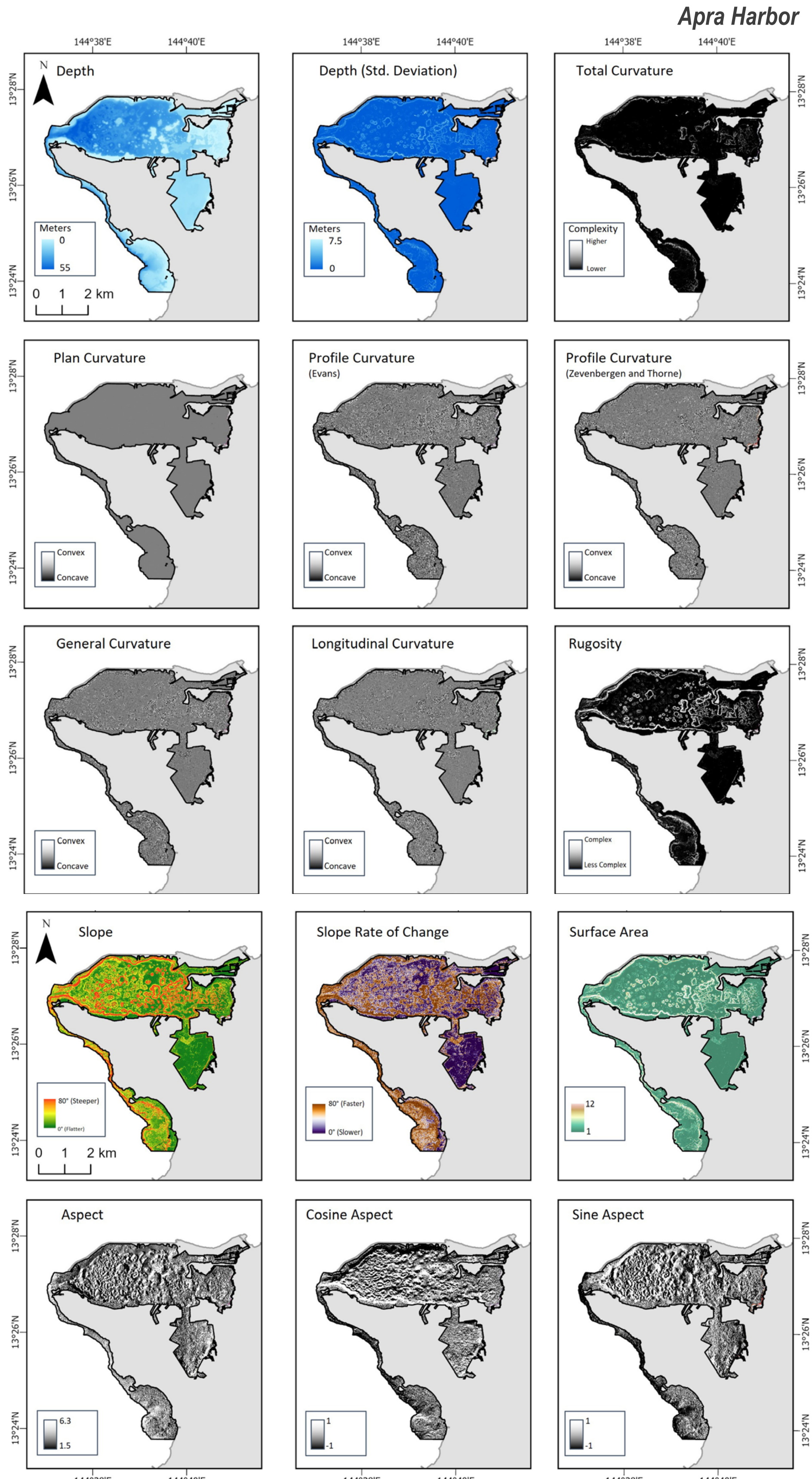


Figure A3. Maps depicting the topographic predictors used to create the habitat predictions for Apra Harbor.

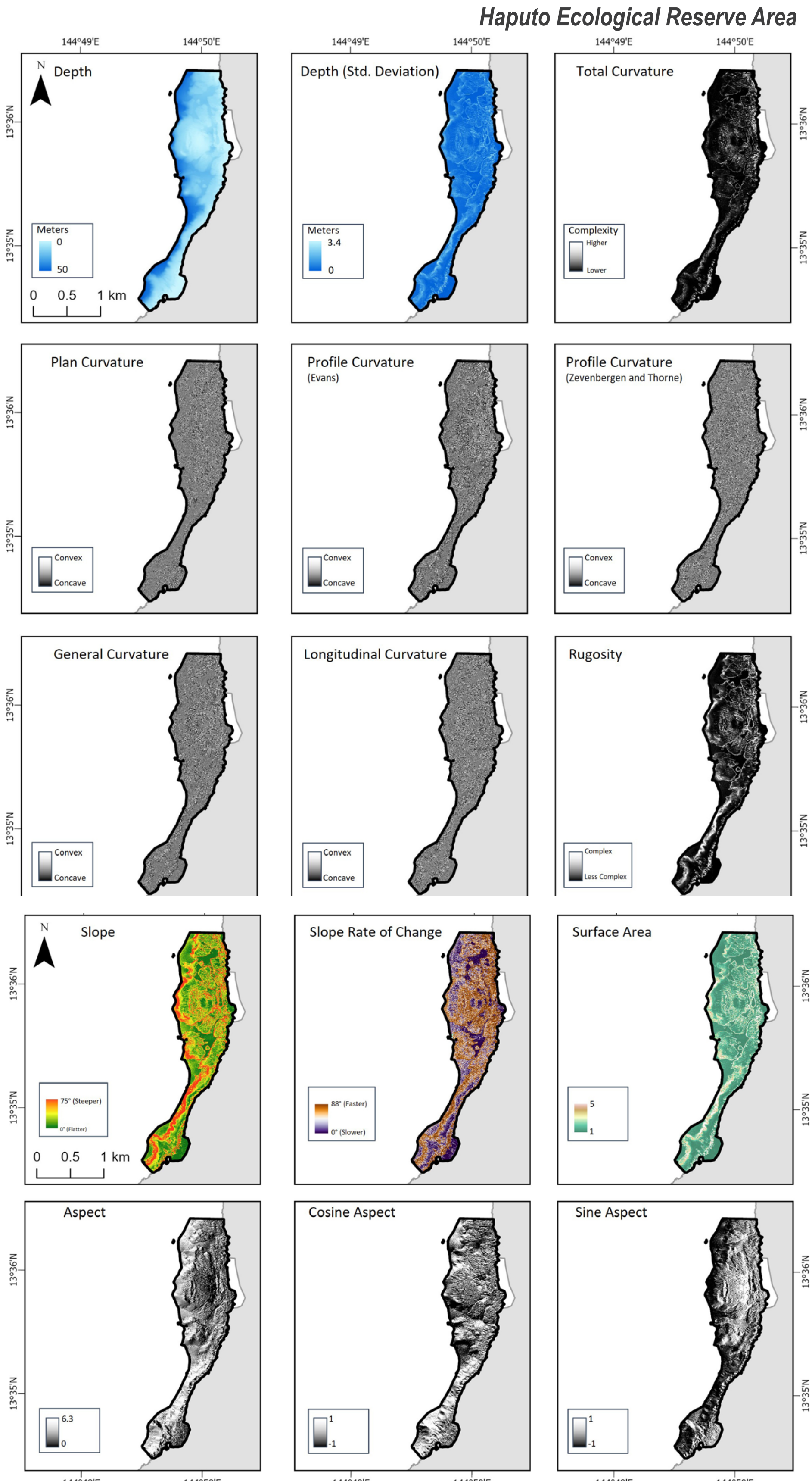


Figure A4. Maps depicting the topographic predictors used to create the habitat predictions for Haputo ERA.

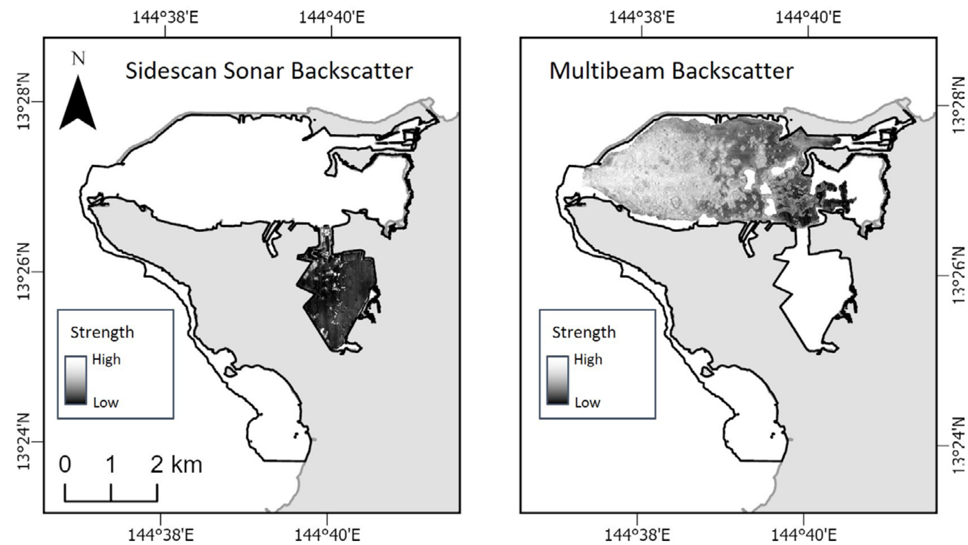


Figure A5. Map depicting the sidescan (left) and multibeam (right) backscatter used to create the habitat predictions for Apra Harbor.

Apra Harbor

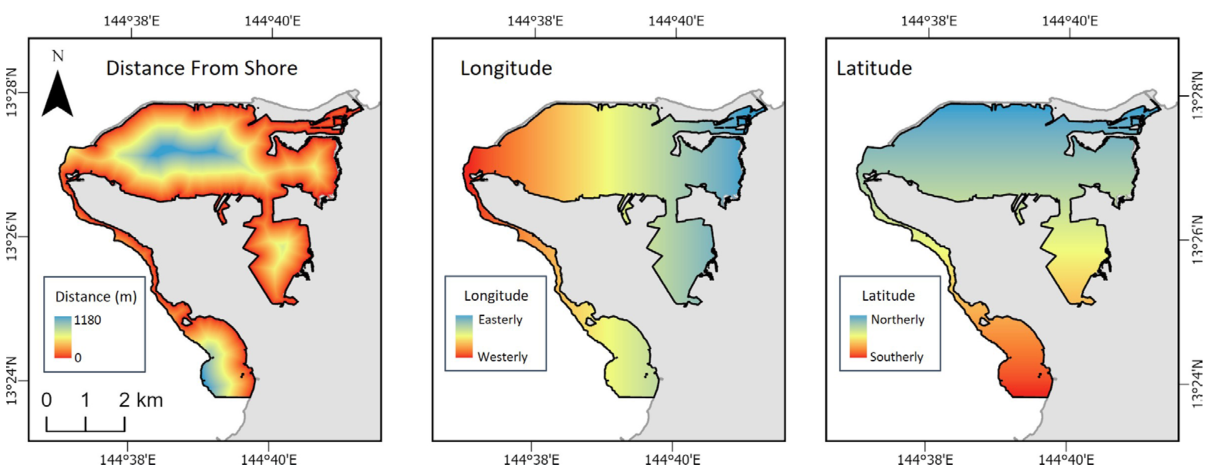


Figure A6. Maps depicting the geographic predictors used to create the habitat predictions for Apra Harbor.

Haputo Ecological Reserve Area

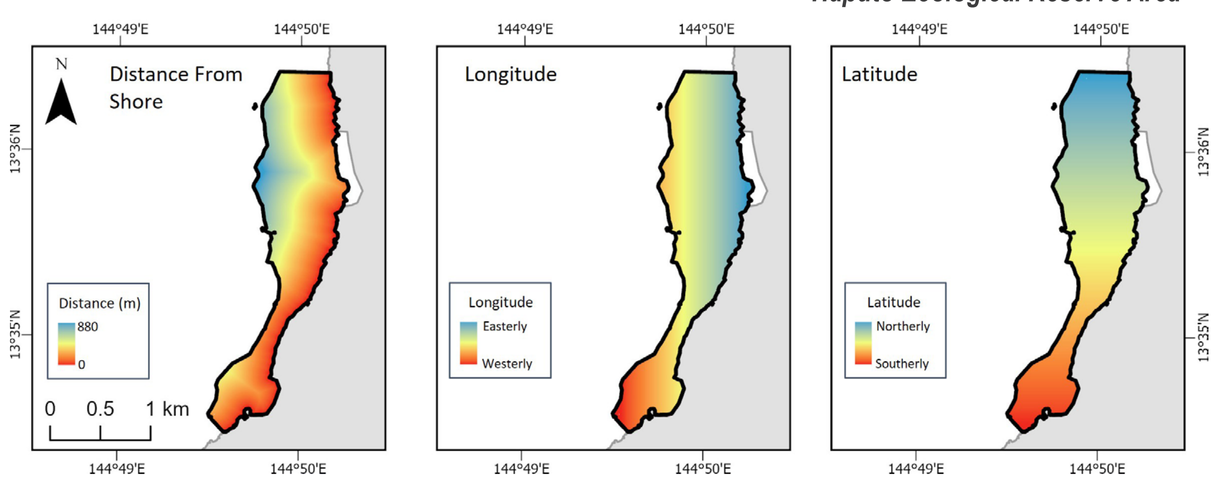


Figure A7. Maps depicting the geographic predictors used to create the habitat predictions for Haputo ERA.

Appendix B

Figure B1. Map showing location of feedback provided during expert review of habitat predictions and map in Apra Harbor. No comments were made about habitat predictions or maps in Haputo ERA. The numbers in the map correspond to the numbers in Table B1.

Table B1. Feedback from and responses to expert review of habitat predictions and map in Apra Harbor. No comments were made about habitat predictions or maps in Haputo ERA.

#	Comment	Response
1	I'm surprised that this doesn't have a higher probability for *Porites rus*; as I recall, that's what makes up most of this reef. This area shows a high probability of Live Coral Reef (All Species), but not for *Porites rus*. This makes me wonder if there is some factor that reduces the model's ability to predict this species?	You're correct. The main driver of the *Porites rus* model (slide 2) is total curvature (slide 3). In that location, total curvature is low compared with surrounding reef (slides 4 & 5), which may explain why probabilities are lower than expected. This is a limitation of the bathy data resolution and, therefore, a limitation of the prediction. We'll plan to discuss/highlight this area in the report in the use/limitations section.
2	From the areas I am familiar with, the maps look fairly accurate within the harbor. Outside of the harbor, along Orote point, I am surprised by the high probability of occurrence for branching (but not surprised by encrusting) corals probability of occurrence. Overall I think the map looks great, and I look forward to using this map to inform our reef restoration activities!	That's great to hear the maps look reasonably accurate! About branching corals along Orote point, we found multiple locations with low % cover of branching corals (slide 6), which is why the model predicts higher probabilities. For this exercise, branching corals included A*cropora, Pocillopora,* and *Porites*, and were marked present if they had ≥1% cover. At the end of this project, we plan to make the georeferenced DLSR underwater photos available publicly so that folks can see for themselves what's on the seafloor at each location. The goal is to make all the data publicly available by March/April.

Appendix C

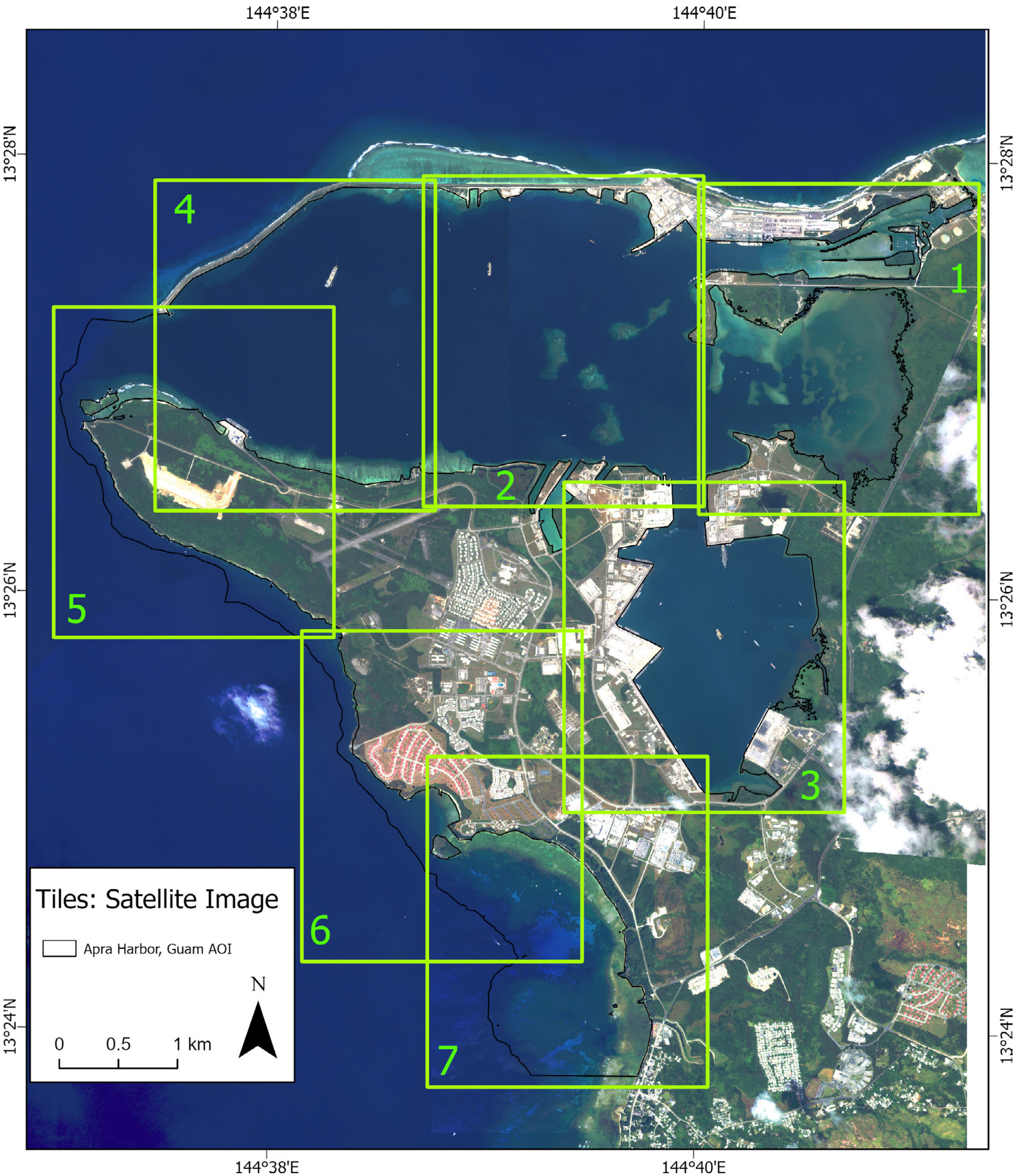


Figure C1. Map showing the location of tiles 1 to 7 for Apra Harbor. AOI = area of interest.

Figure C2. Tile 1 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

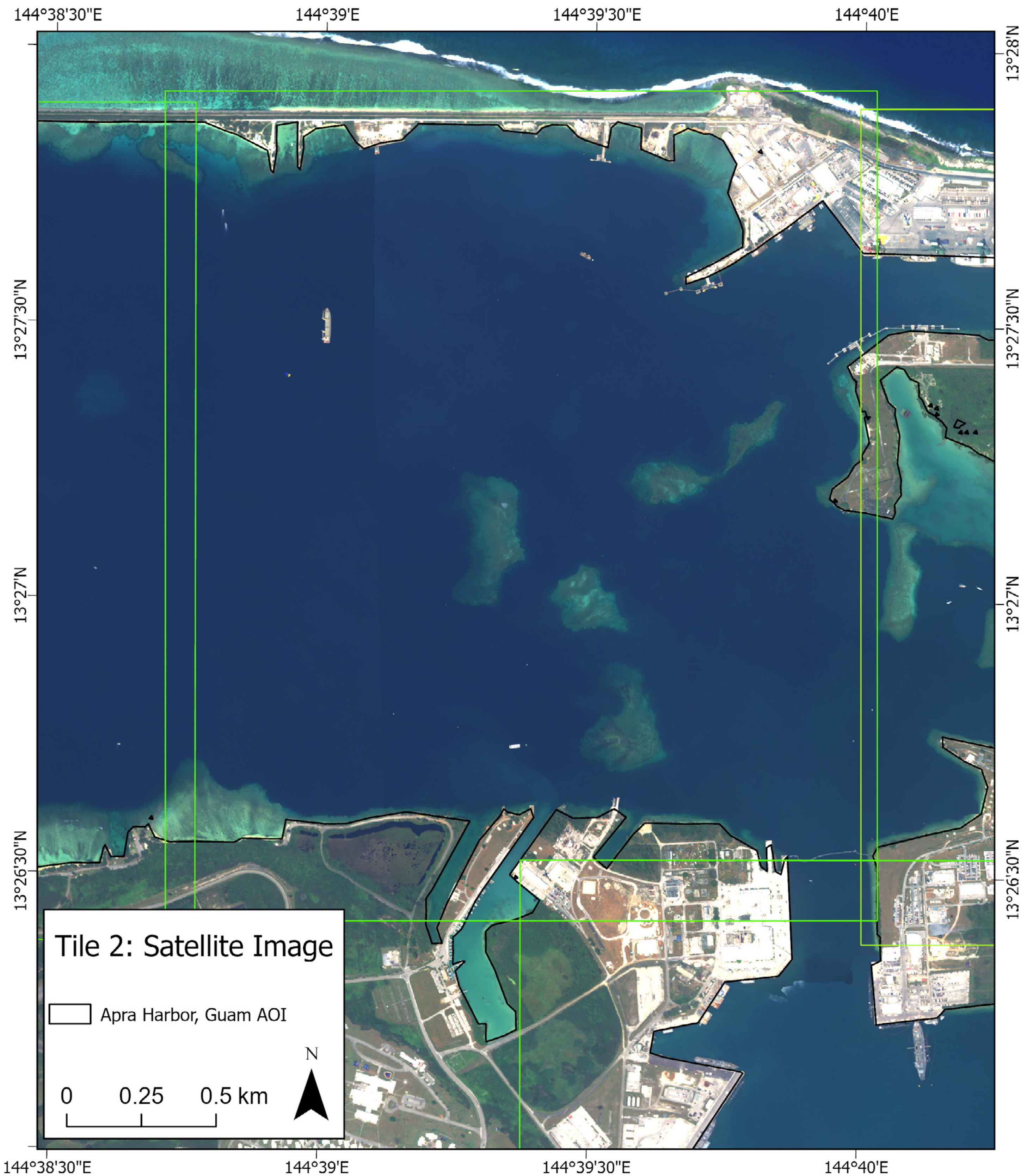


Figure C3. Tile 2 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

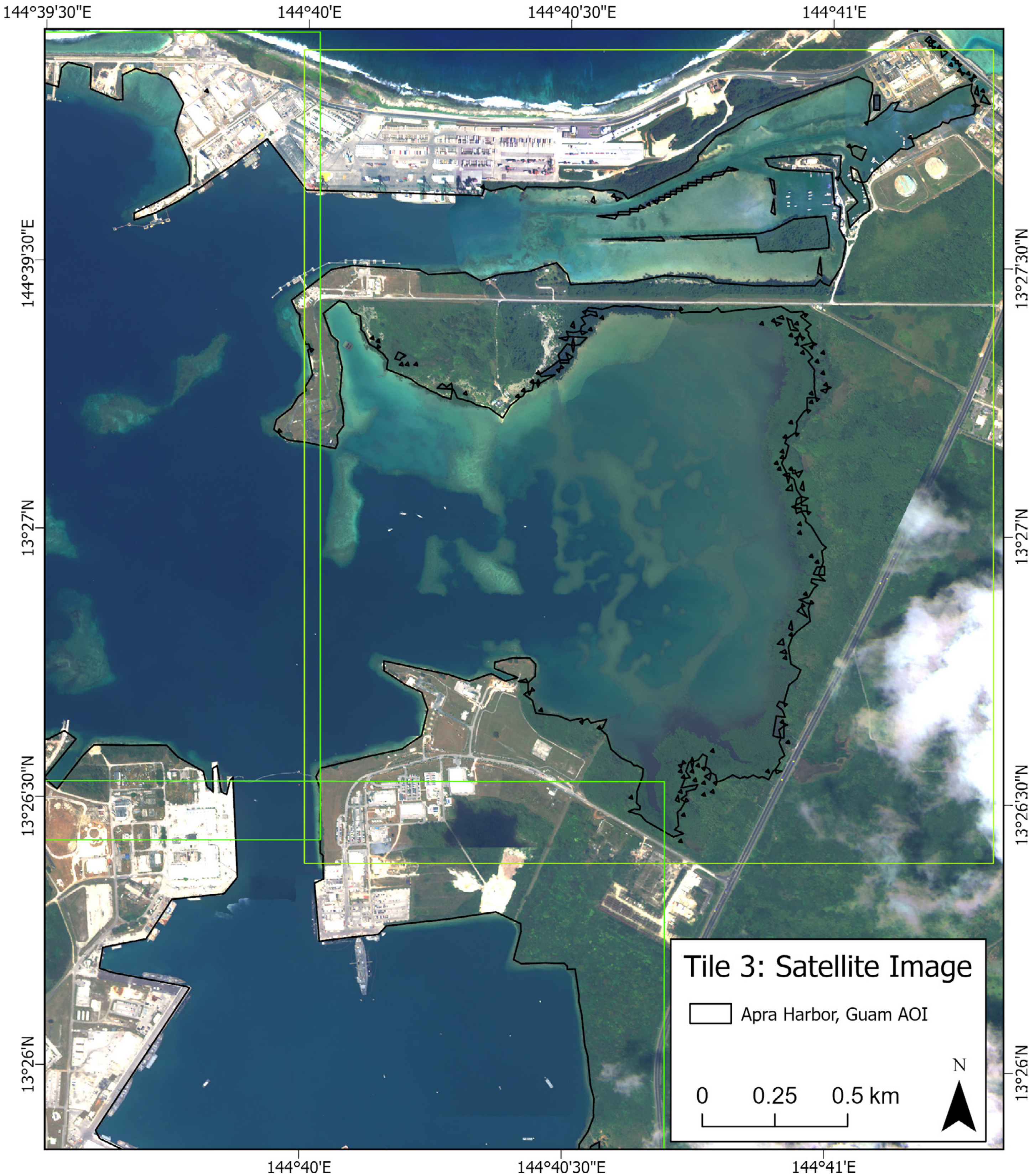


Figure C4. Tile 3 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

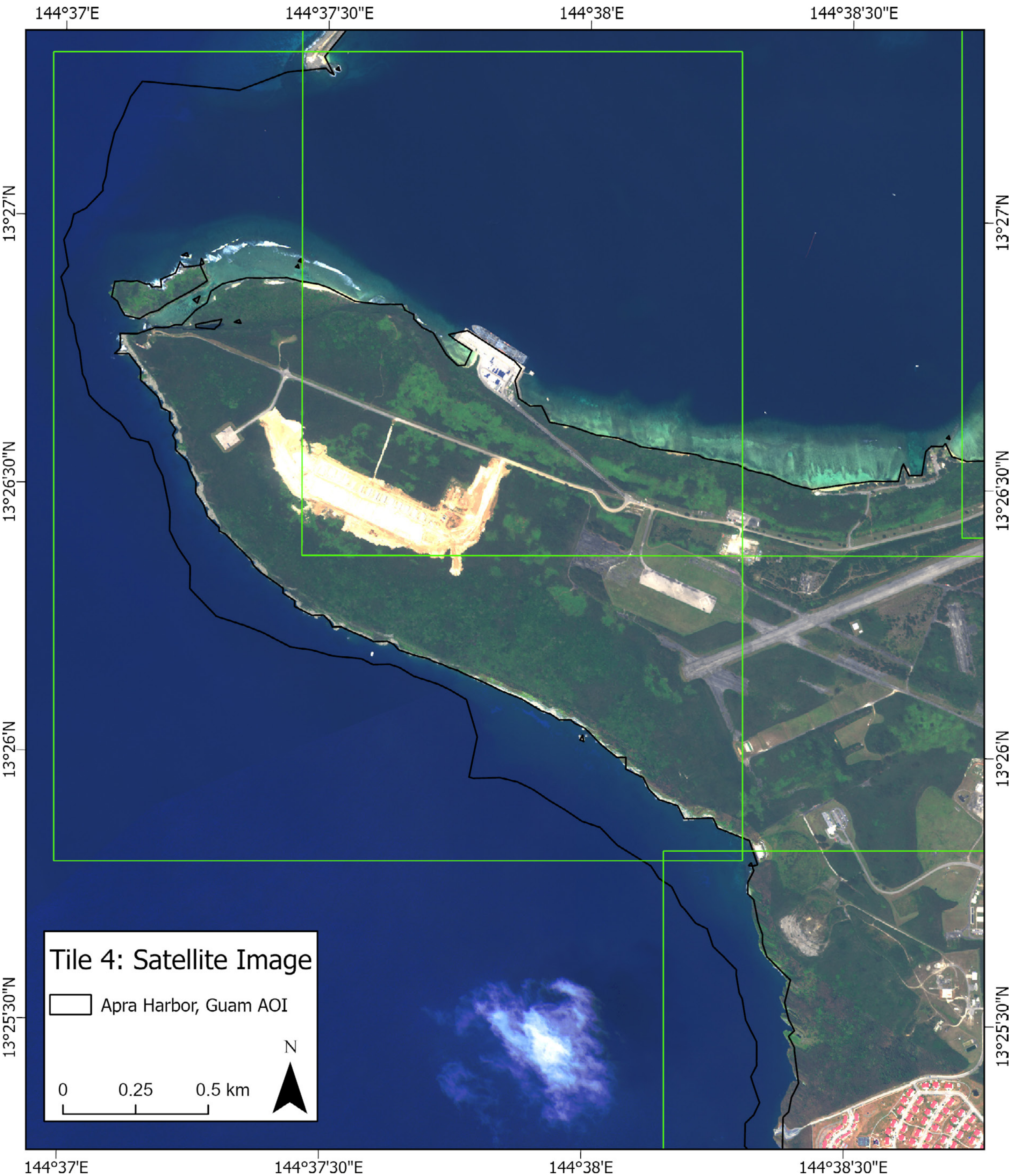


Figure C5. Tile 4 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

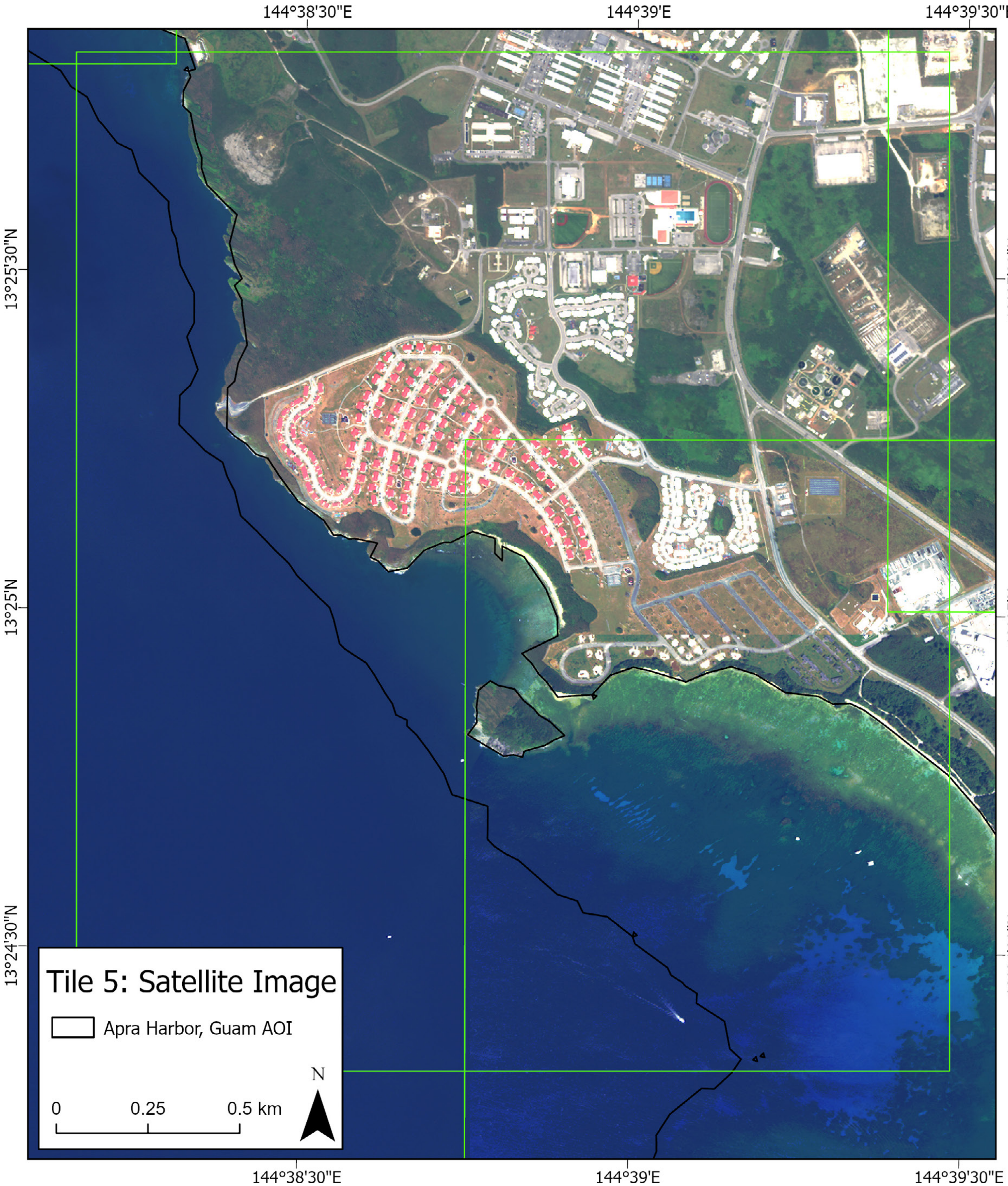


Figure C6. Tile 5 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

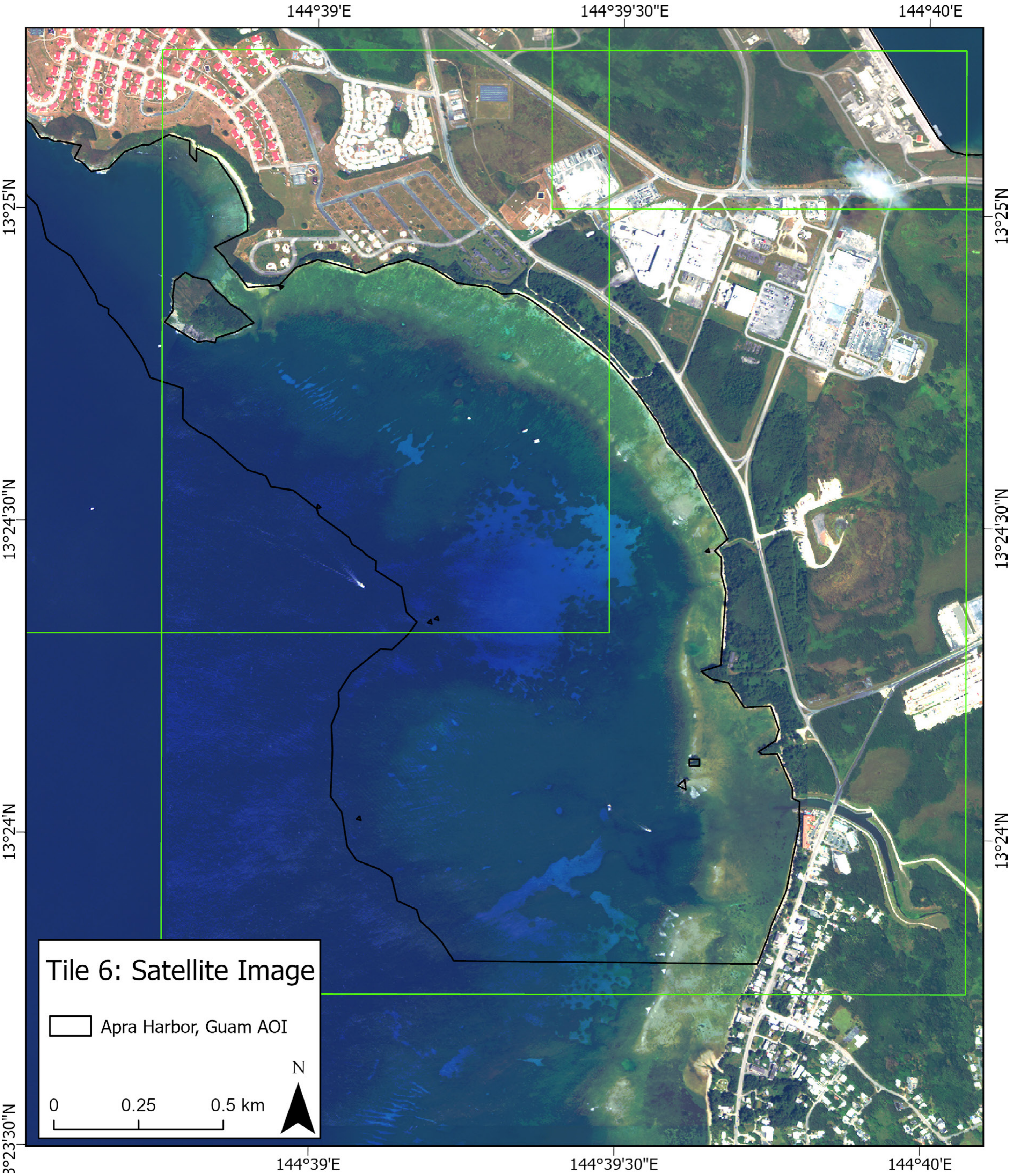


Figure C7. Tile 6 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

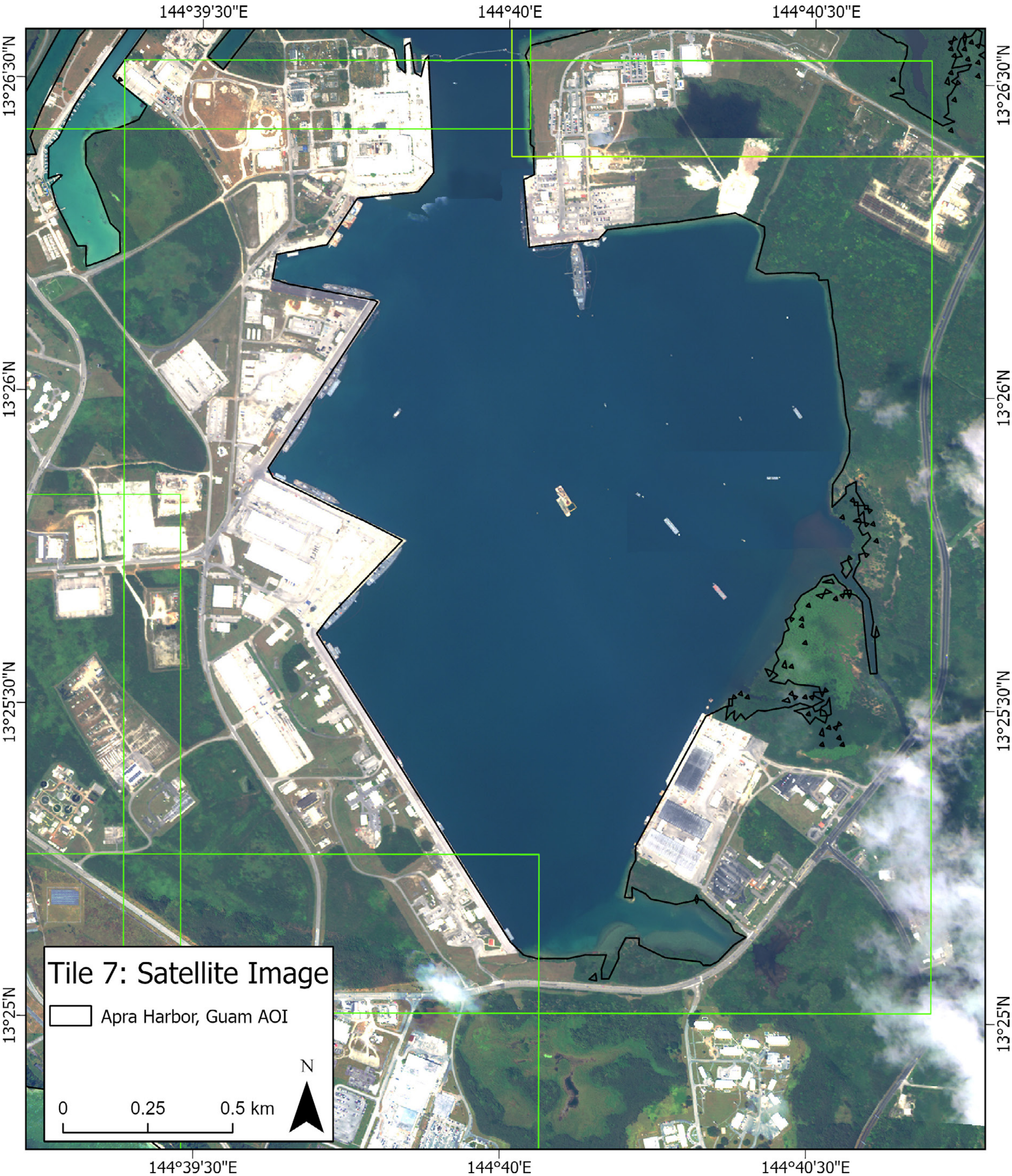


Figure C8. Tile 7 showing the satellite image mosaic used to create benthic habitat maps for Apra Harbor. AOI = area of interest.

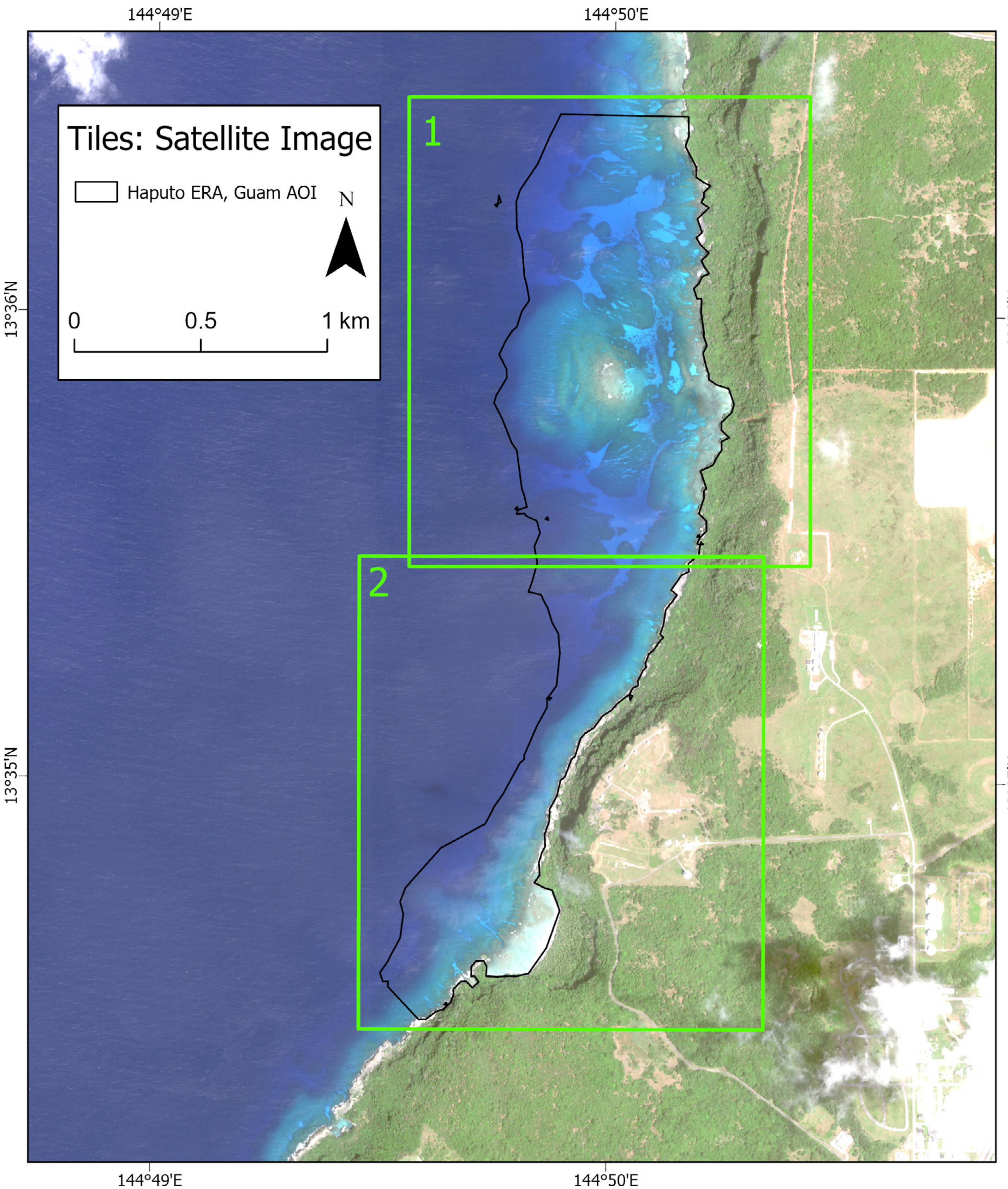


Figure C9. Map showing the location of tiles 1 and 2 for Haputo Ecological Reserve Area. AOI = area of interest.

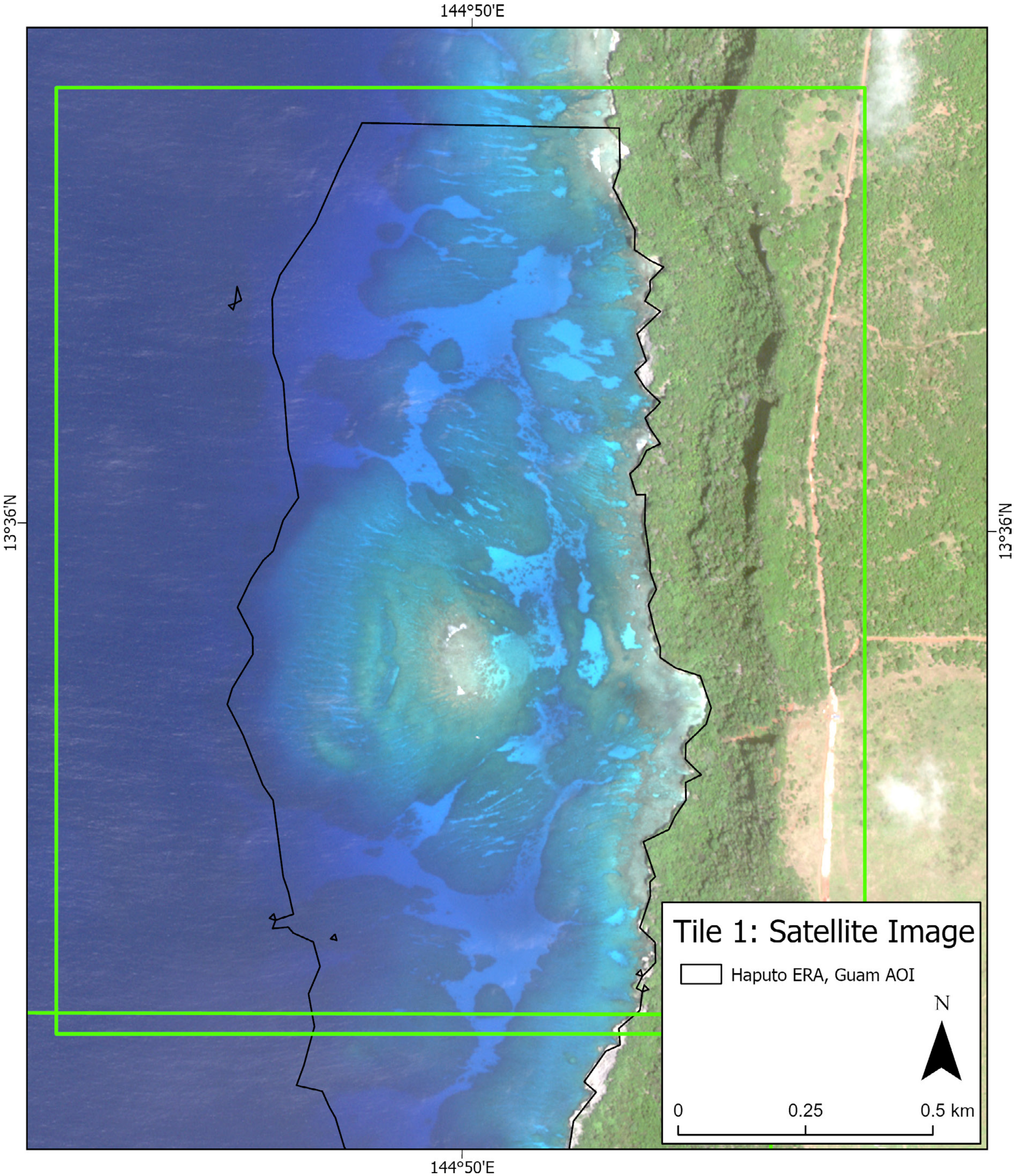


Figure C10. Tile 1 showing the satellite image mosaic used to create benthic habitat maps for Haputo Ecological Reserve Area. AOI = area of interest.

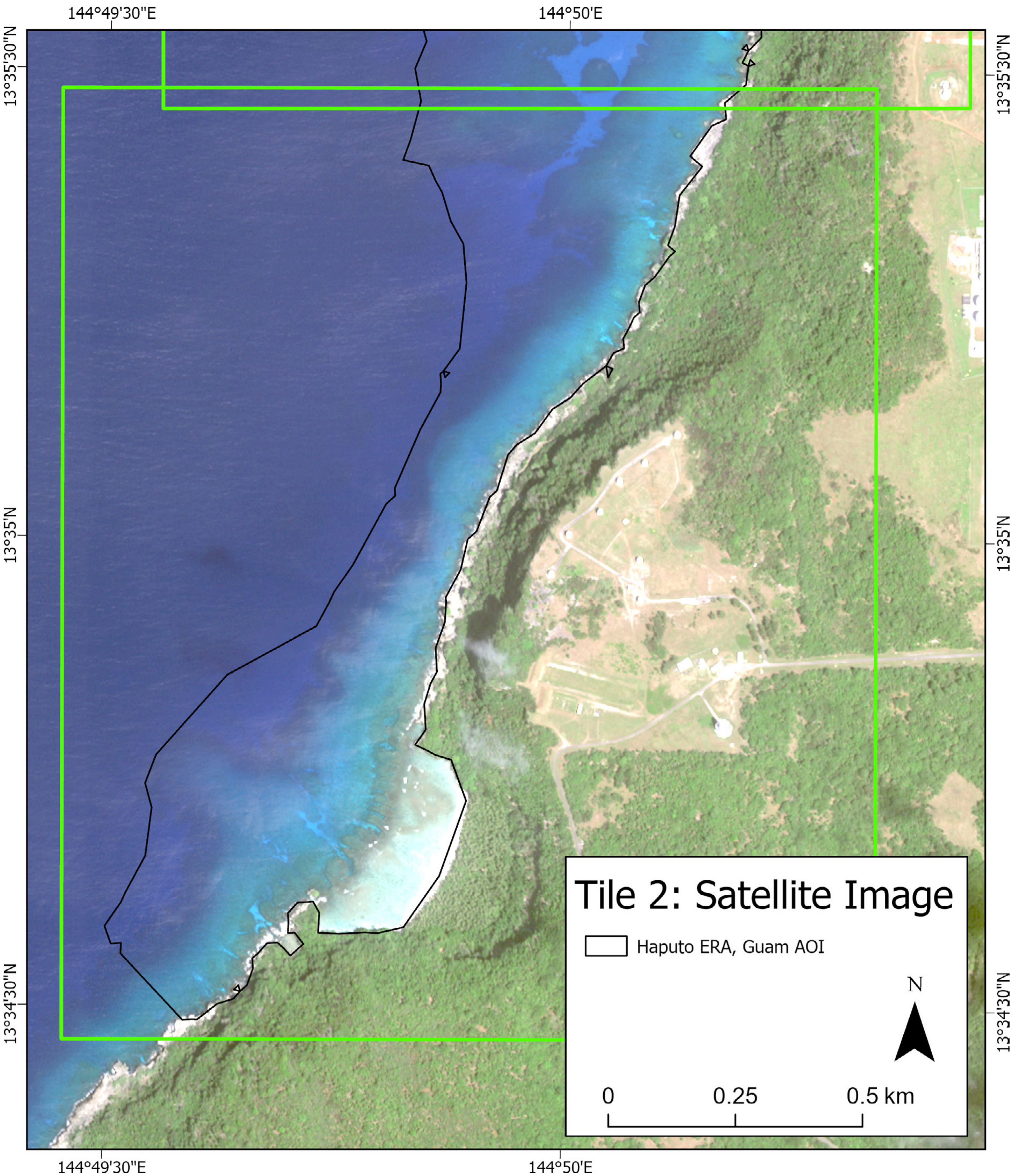


Figure C11. Tile 2 showing the satellite image mosaic used to create benthic habitat maps for Haputo Ecological Reserve Area. AOI = area of interest.

Appendix D

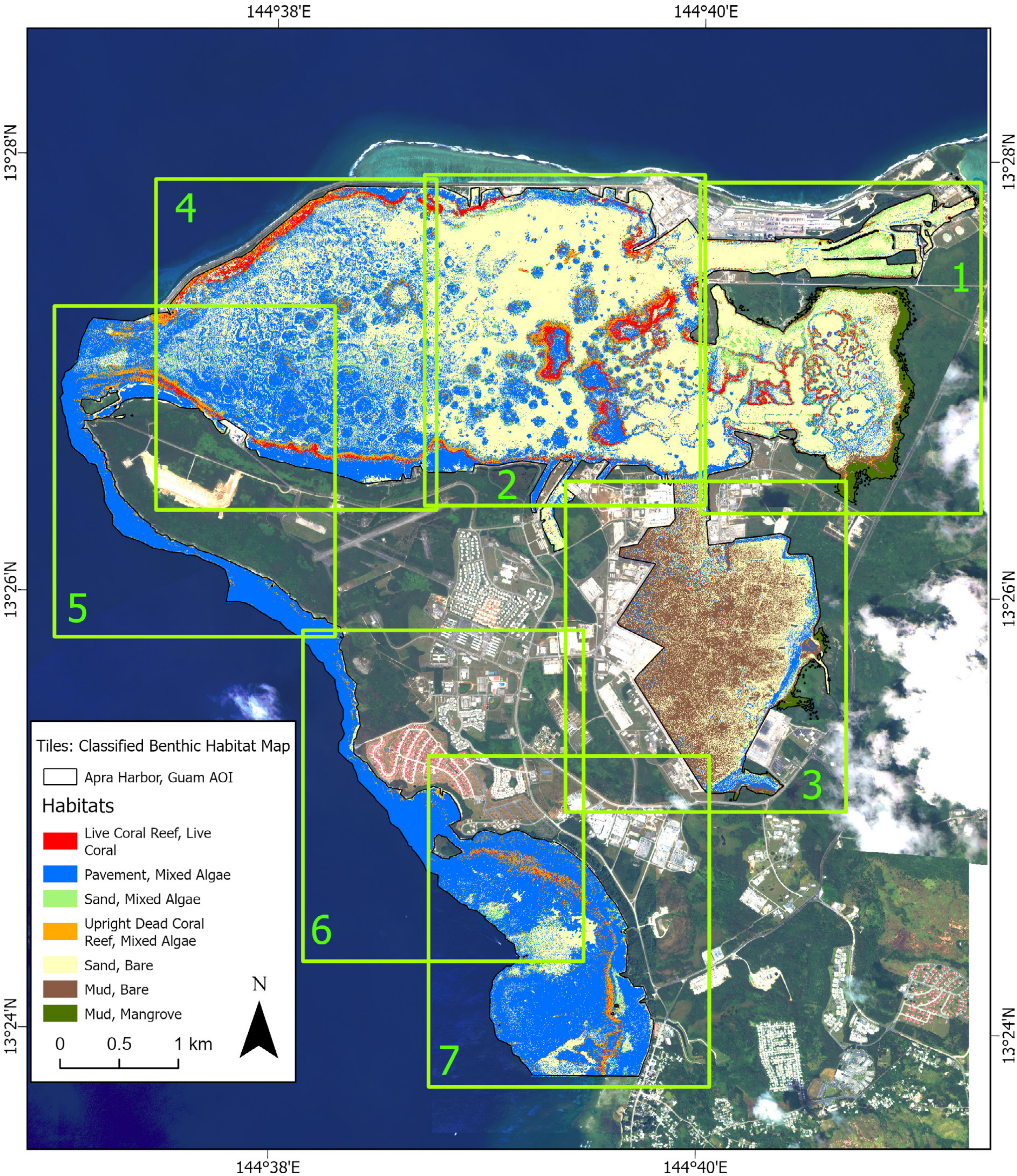


Figure D1. Map showing the location of tiles 1 to 7 for Apra Harbor. AOI = area of interest.

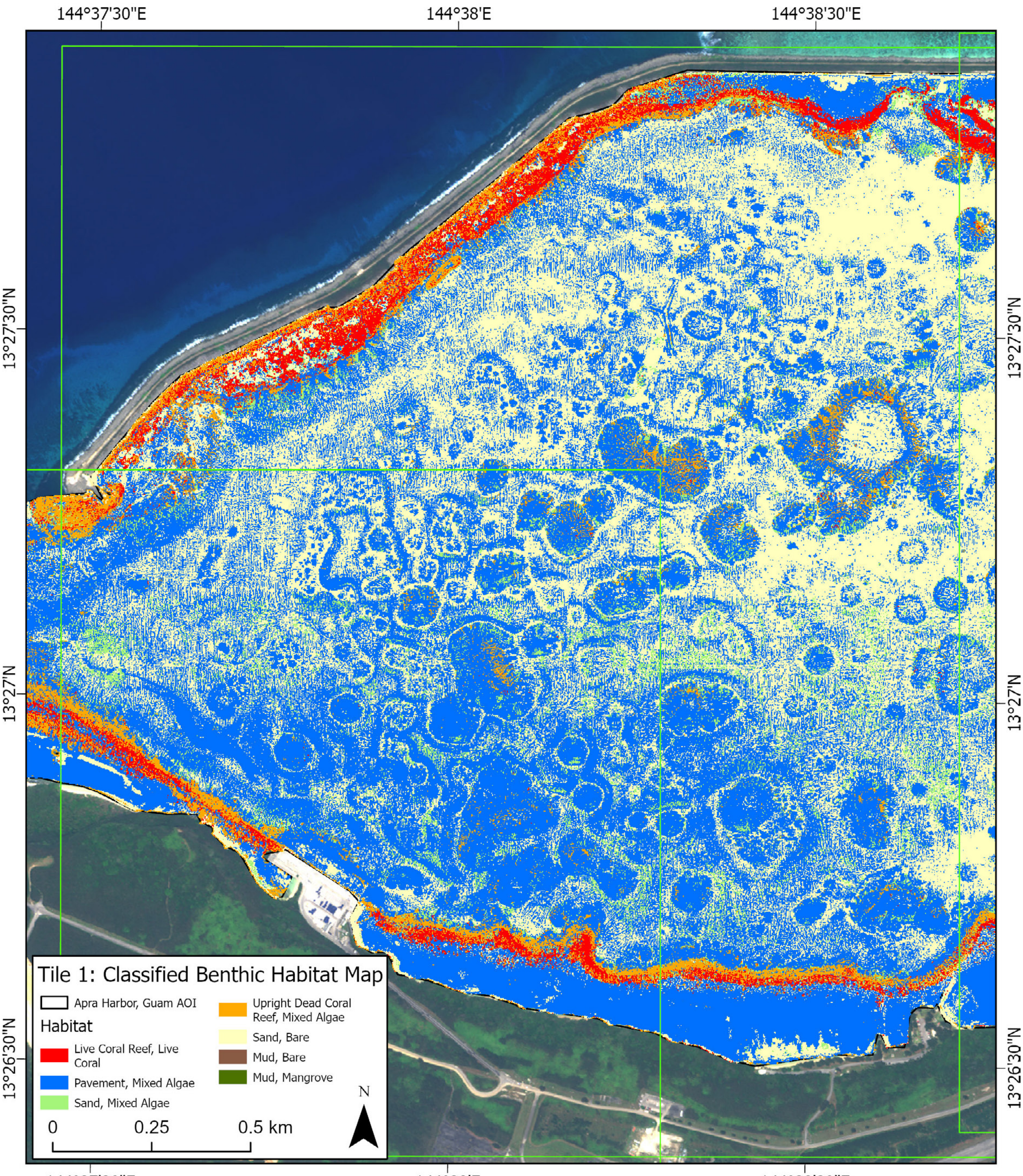


Figure D2. Tile 1 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

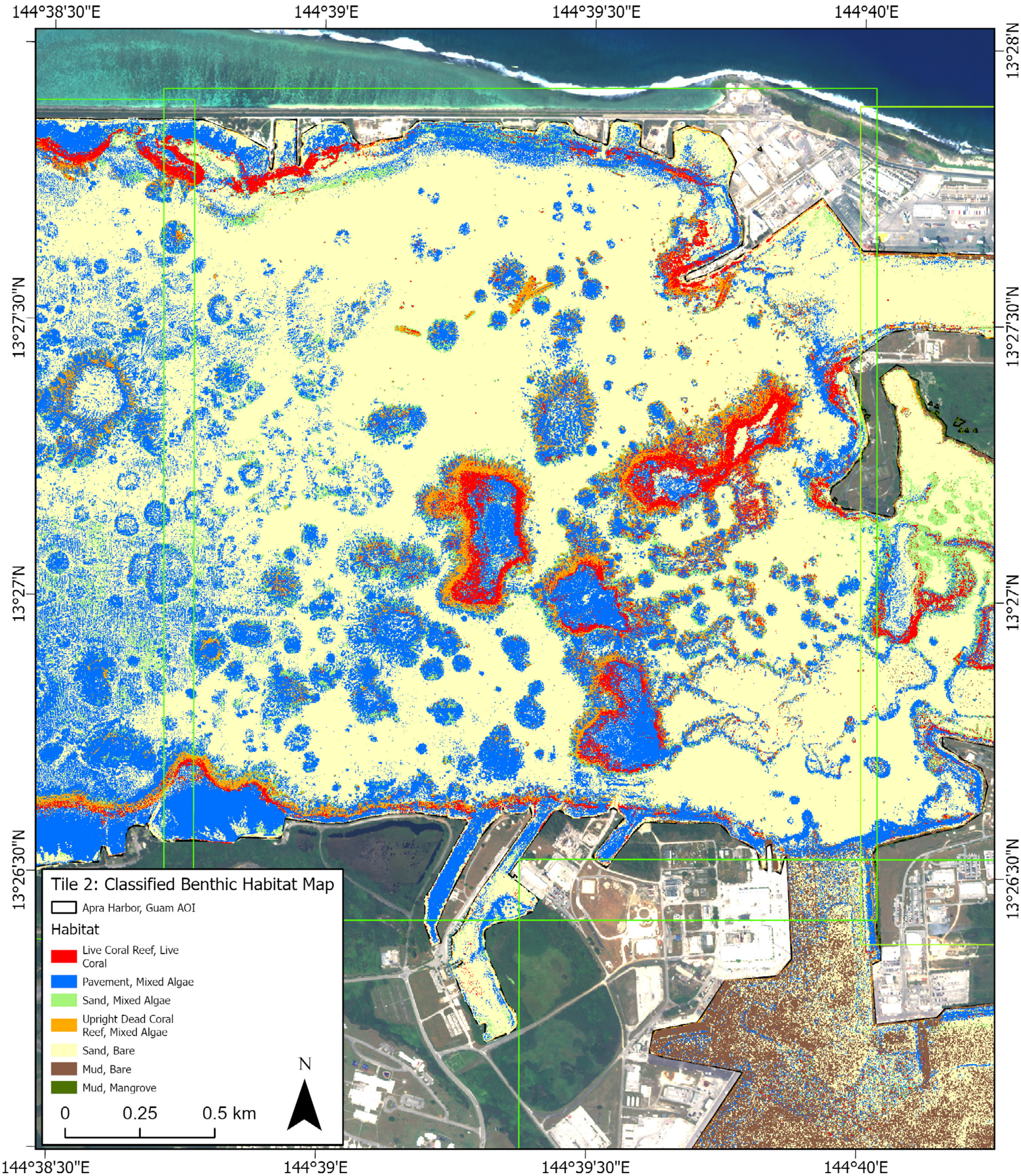


Figure D3. Tile 2 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

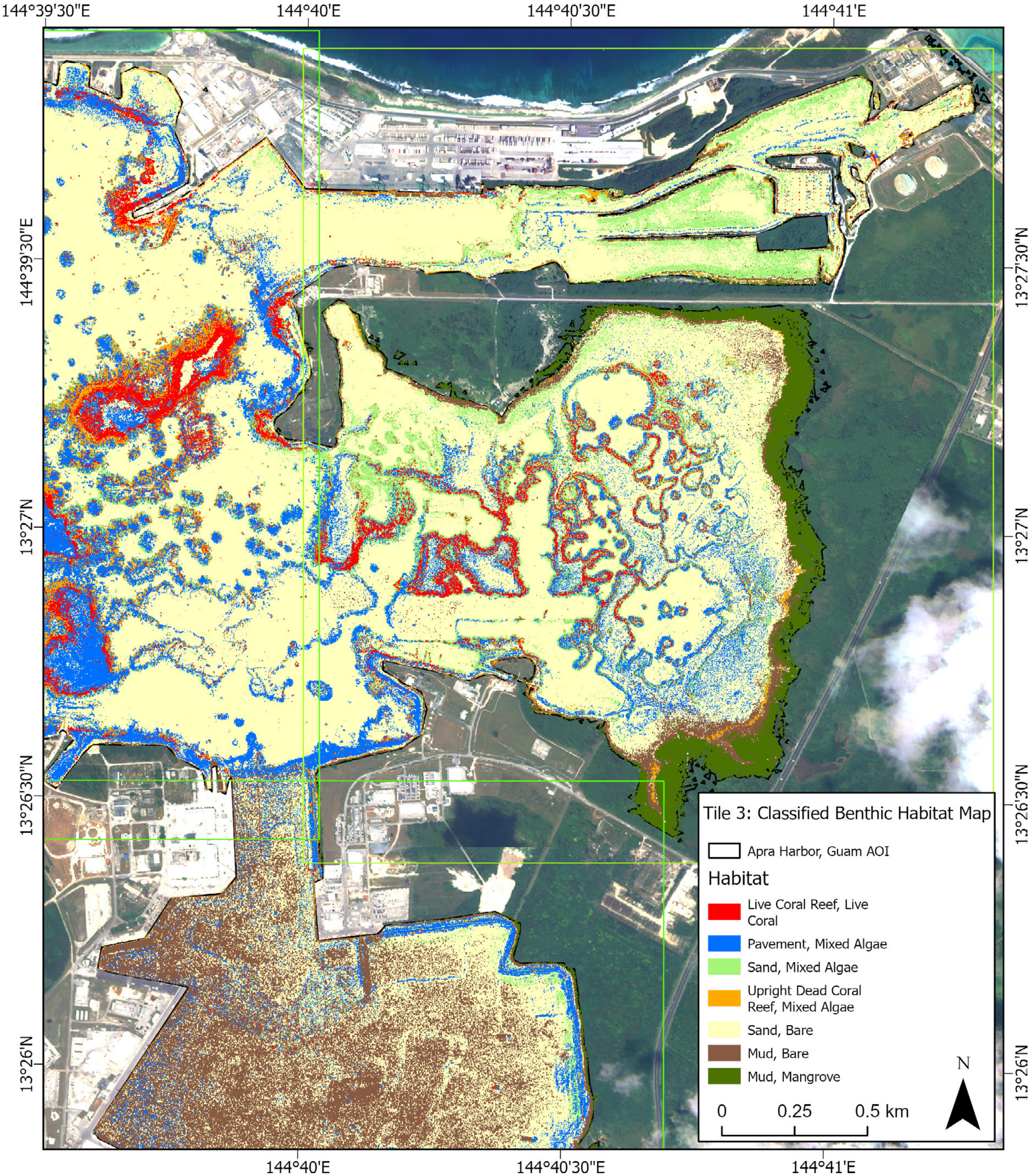


Figure D4. Tile 3 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

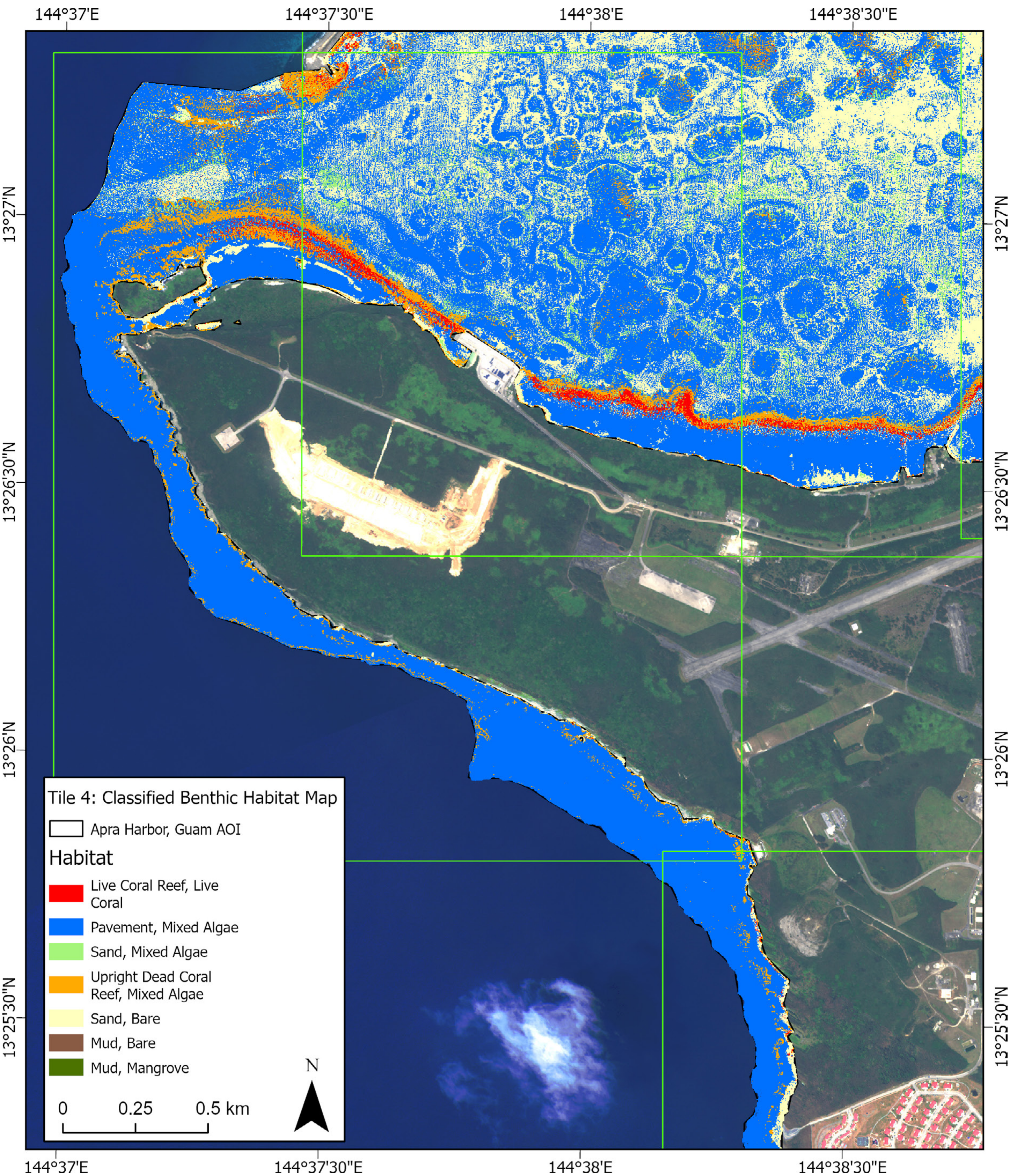


Figure D5. Tile 4 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

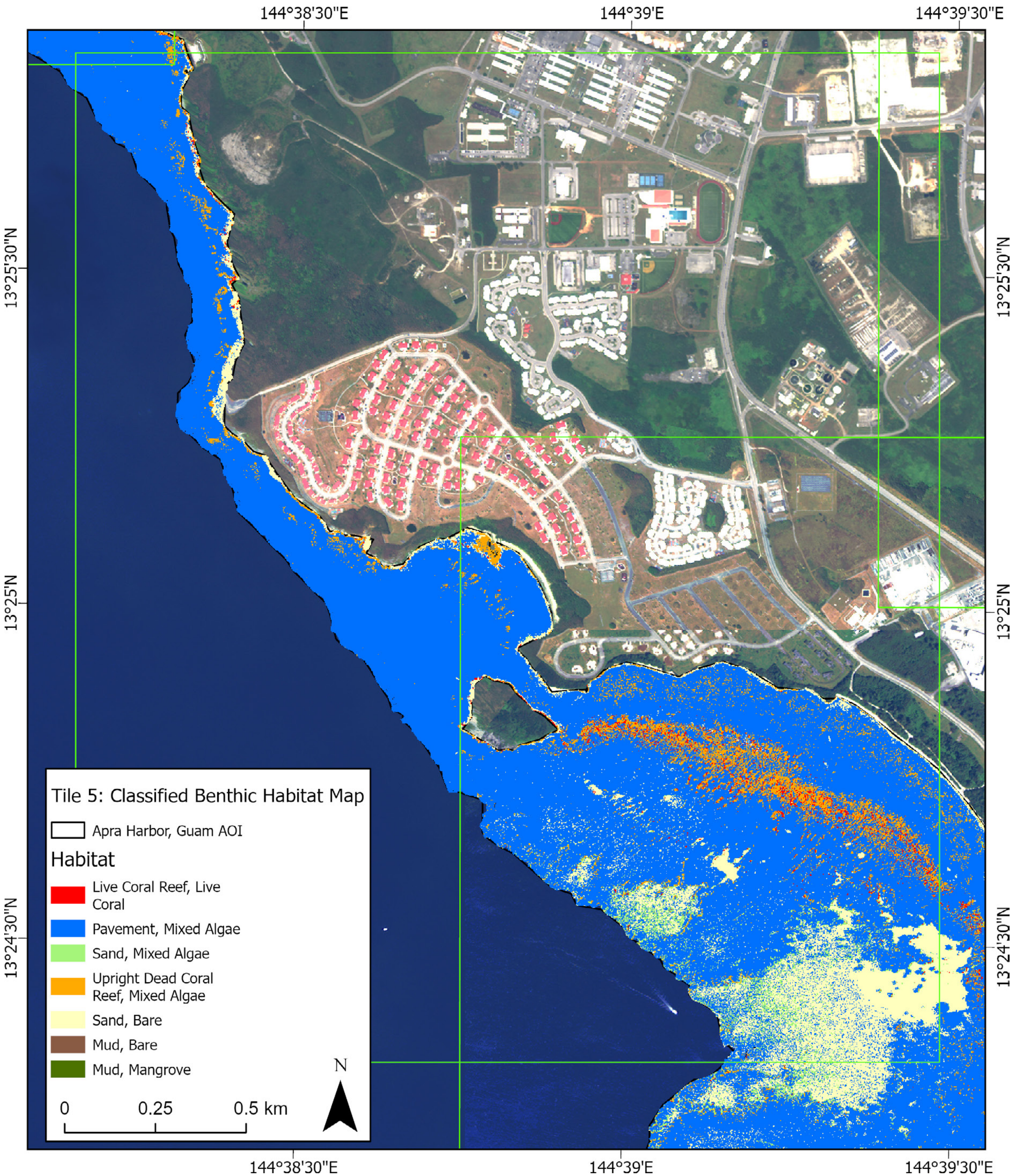


Figure D6. Tile 5 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

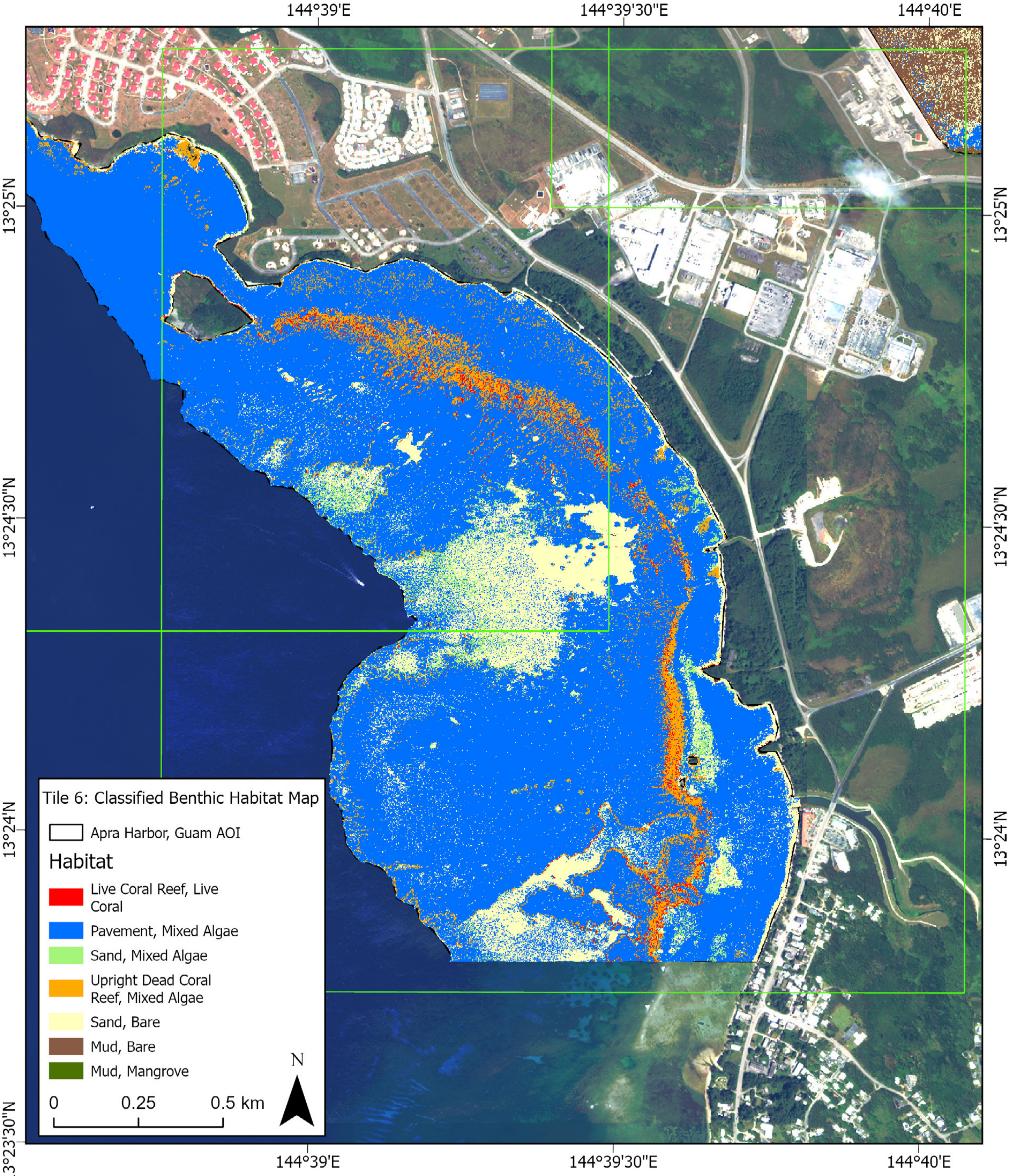


Figure D7. Tile 6 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

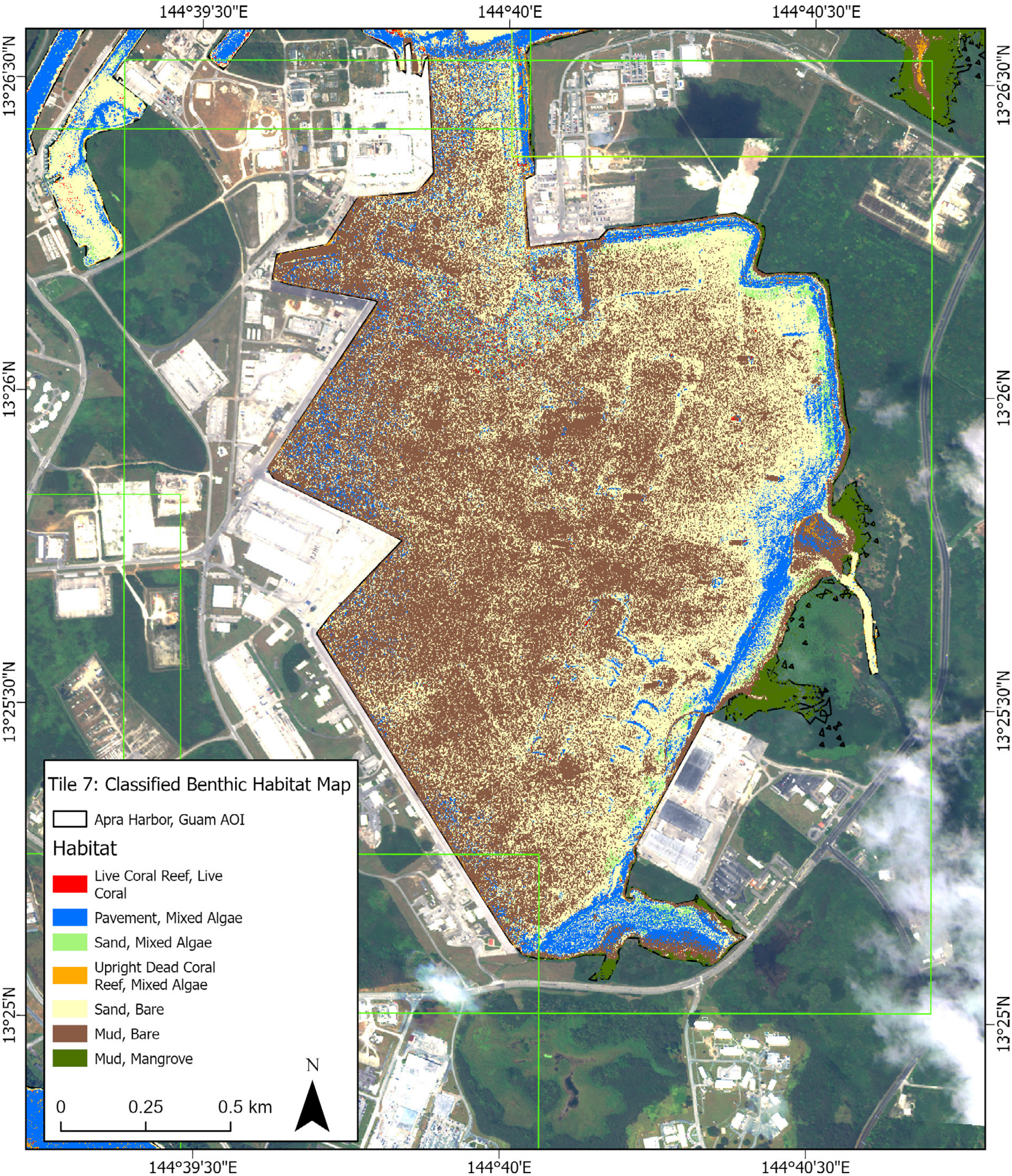


Figure D8. Tile 7 showing the classified benthic habitat map for Apra Harbor. AOI = area of interest.

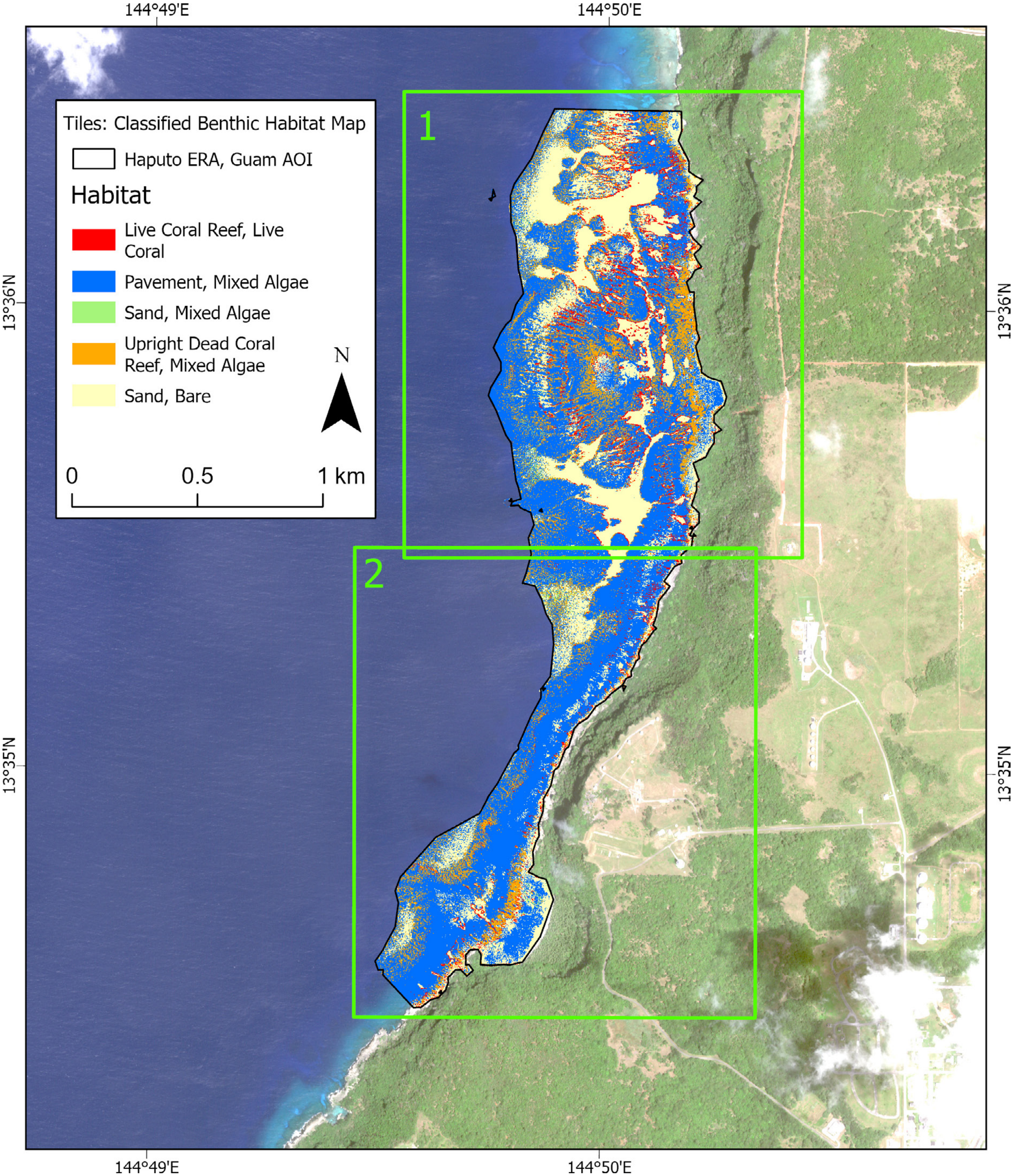


Figure D9. Map showing the location of tiles 1 to 2 for Haputo Ecological Reserve Area. AOI = area of interest.

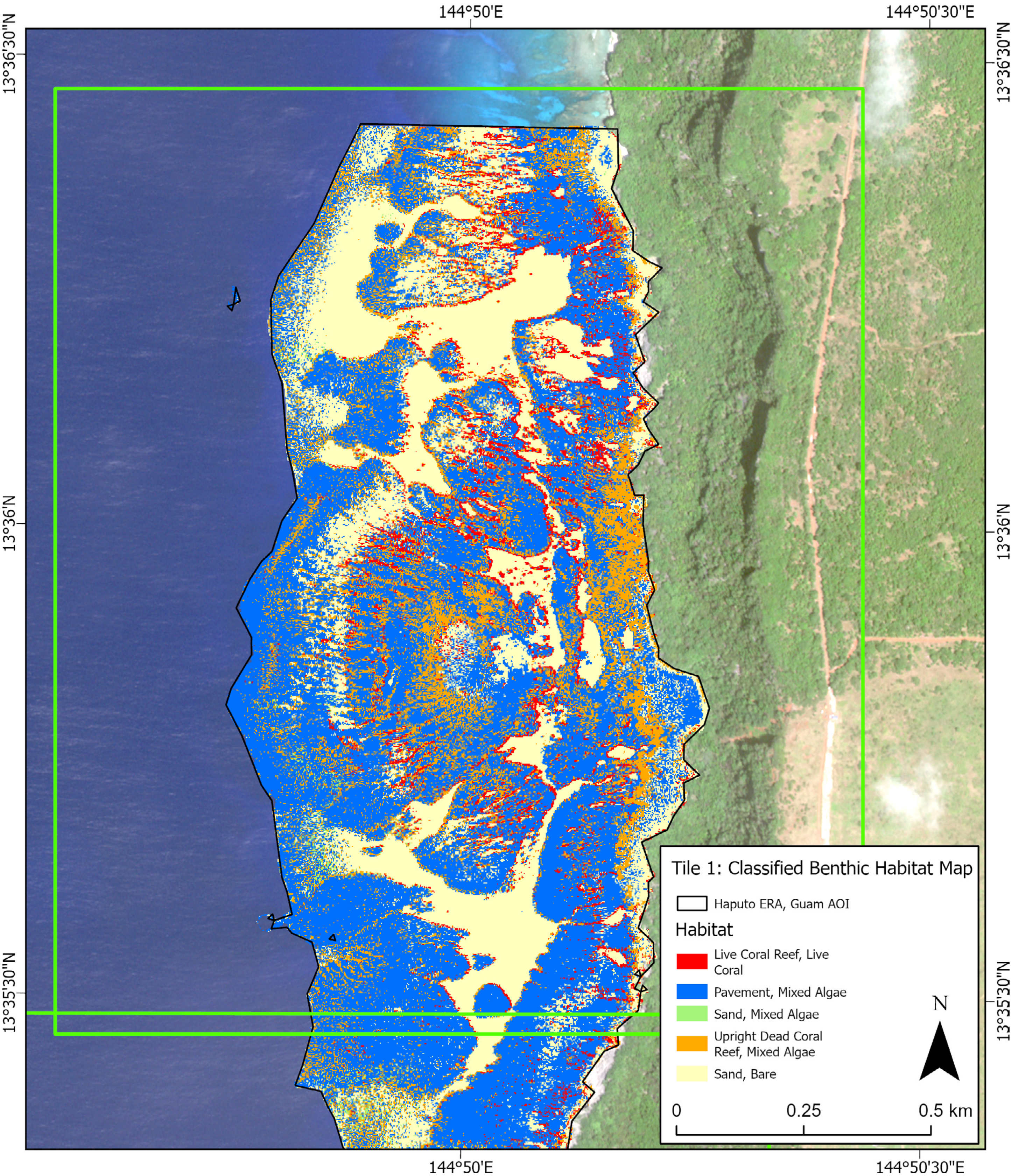


Figure D10. Tile 1 showing the classified benthic habitat map for Haputo Ecological Reserve Area. AOI = area of interest.

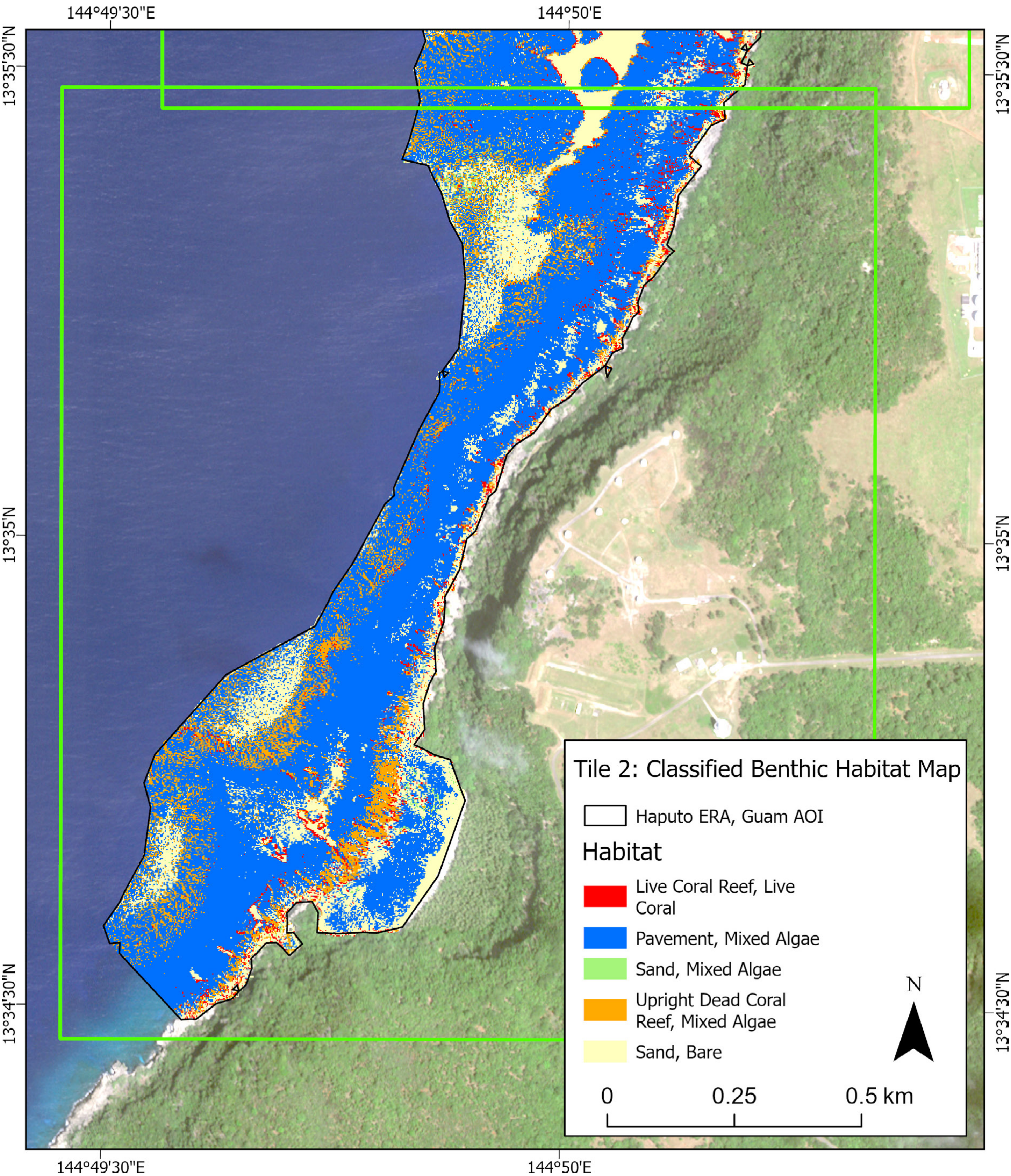


Figure D11. Tile 2 showing the classified benthic habitat map for Haputo Ecological Reserve Area. AOI = area of interest.

U.S. Department of Commerce

Gina M. Raimondo, *Secretary*

National Oceanic and Atmospheric Administration

Richard Spinrad, *Under Secretary for Oceans and Atmosphere*

National Ocean Service

Nicole LeBoeuf, *Assistant Administrator for National Ocean Service*