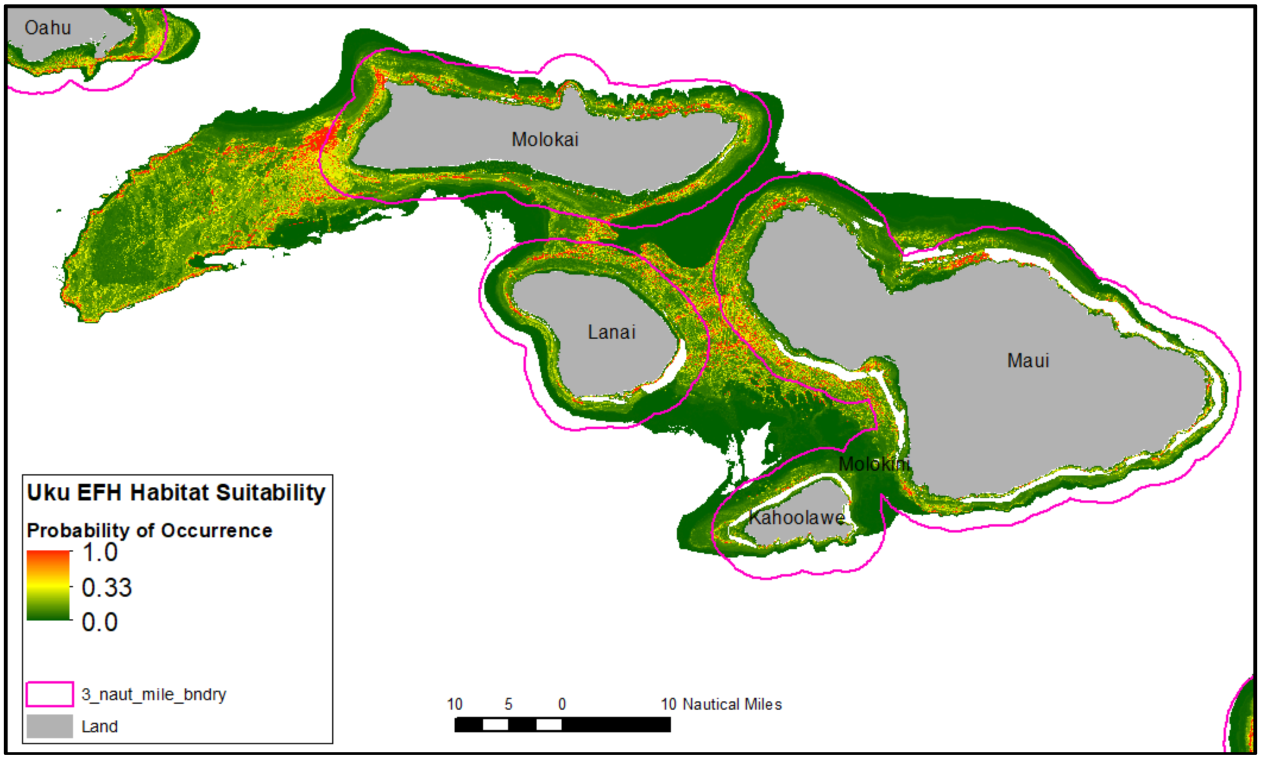
Model-based Essential Fish Habitat Definitions for the Uku *Aprion virescens* in the Main Hawaiian Islands



by

Erik C. Franklin

for

Western Pacific Regional Fishery Management Council

1164 Bishop Street, 1400 Honolulu, Hawaii 96813

June 30, 2021

This report was created with funding support from the Western Pacific Regional Fishery Management Council under contract 21-ADMIN-##. The document should be cited as follows:

> Franklin, E.C. 2021. Model-based essential fish habitat definitions for the Uku *Aprion virescens* in the Main Hawaiian Islands. Report to WPRFMC. 33 pp.

Acknowledgements: Thanks to Tye Kindinger, Jake Asher, Eric Conklin, Laura Gajdzik, and Audrey Rollo for preparing and sharing data; and Kisei Tanaka and Michael Parke for feedback on the analysis.

Table of Contents

Abstract

The description of essential fish habitat is required for federally managed fishery species. Species distribution models (SDMs) have been utilized in marine science and conservation to identify and predict suitable habitats for many species. SDMs can involve a range of models that vary in structure and complexity. Boosted regression tree models were fit to fishery-independent data to define the geographic extent of essential fish habitat (EFH) of sub-adult/adult life stages of a federally managed snapper species, *Aprion virescens*, commonly called "Uku", in the main Hawaiian Islands. Due to differences in survey data collection methods, separate SDM models for Uku were constructed for shallow (0-30 m depth) depths using NOAA fish diver surveys and deeper (30-300 m) depths using NOAA and UH baited stereo-video camera arrays. For shallow models, aspect (i.e., direction that habitat slope faces), depth, and wave heights were strong predictors of Uku occurrence, while depth was the predominant habitat variable for the deep model. Output from the SDMs were used to create maps delineating Uku EFH including continuous probability of occurrence maps as well as categorical maps showing EFH "hot spots", "core habitats", and "basic EFH", based on 25%, 50%, and 95% quantiles of predicted occurrence, respectively. For shallow depths, Uku hot spots were 0% of the MHI, core habitat was 0.2%, basic EFH was 55.4%, and other habitat was 44.4%. For deep depths, Uku hot spots were 0.09% of the MHI, core habitat was 2.4%, basic EFH was 59.8%, and other habitat was 37.3%. The maps are a visual display of the predicted quantitative relationship between Uku and their habitat and can be used to inform marine conservation and management activities in the main Hawaiian Islands. These analyses represent the first model-based approach to delineating EFH for a U.S. Western Pacific stock and could serve as a framework for the EFH descriptions of other managed species in the region. A future phase of this work will expand the environmental covariates in the SDMs to include dynamic variables such as seawater temperature, salinity, productivity, and current velocity.

INTRODUCTION

The description and identification of Essential Fish Habitat (EFH) for all species managed under a US fisheries management plan is mandated by the Magnuson-Stevens Fishery Conservation and Management Act (1996). EFH is generally defined therein as "those waters and substrate necessary for fish for spawning, breeding and growth to maturity". The National Marine Fisheries Service has provided guidelines to evaluate the quality of available data used to identify EFH should be rated (50 CFR Pt. 600.815):

Level 1: All that is known is where a species occurs based on distribution data for all or part of the geographic range of the species.
Level 2: Data on habitat-related densities or relative abundance of the species are available.
Level 3: Data on growth, reproduction, or survival rates within habitats are available.
Level 4: Production rates by habitat are available.

In the Hawaiian Archipelago, most EFH definitions for managed species have been primarily qualitative descriptions of the distribution and depth range of the observed life stages (i.e., typically adults, but can include juveniles). For the Green Jobfish, *Aprion virescens* Valenciennes 1830 (Figure 1), or "Uku" as it is known locally in Hawaiʻi, the EFH designation was classified as part of the "Shallow Complex" of bottomfish in the Hawaiian Archipelago by the Western Pacific Regional Fishery Management Council (WPRFMC 2016). In the Western Pacific region, four life stage categories are defined for EFH: egg, post-hatch pelagic, post-settlement, and sub-adult/adult. For the sub-adult/adult stage of Uku, the EFH was described as "the benthopelagic zone, including all bottom habitats, in depths from the surface to 240 m bounded by the official US baseline and 240 m isobath" and maps to reflect that description were generated. Definitions for the Uku egg, post-hatch pelagic, and post-settlement stages interpreted in aggregate generically identified all water column (out to 50 mi from shore for eggs) and benthic habitats from 0 to 240 m depths as EFH.

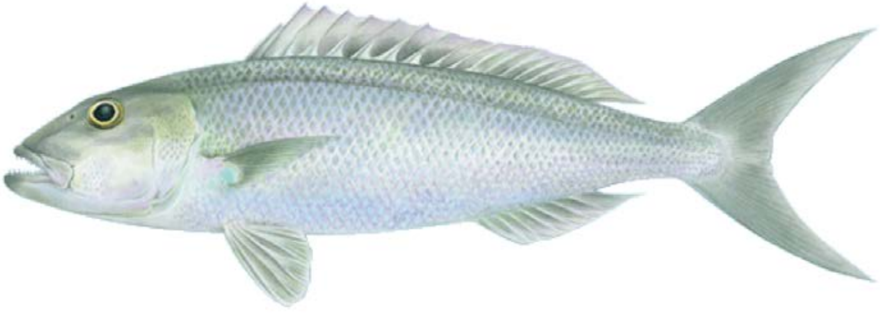


Figure 1. Drawing of adult Green Jobfish or "Uku", *Aprion virescens*. Source: Hawaiʻi DLNR.

This scope of work outlines methods to utilize existing fisheries-independent survey data with geographically-explicit marine habitat information (i.e., GIS data) to identify EFH for the sub-adult/adult phases for *Aprion virescens* in the main Hawaiian Islands (MHI) of the Hawaiian Archipelago (i.e., the Northwestern Hawaiian Islands are not included in this analyses). This work uses statistical and geospatial methods whose complexity depended upon the quality, type, and amount of data available for each species. This work did not attempt to delineate EFH for the egg of post-hatch pelagic stages of the fish species.

To expand on the existing EFH definitions from Uku in the MHI, the purpose of this work is to generate species distribution models (SDMs) for Uku for sub-adult/adult life stages to the highest EFH Levels possible based on available data. In this report, the SDMs were constructed by fitting boosted regression tree (BRT) models to fishery-independent survey data from NOAA and the University of Hawaii. The BRT models provide optimal fits between habitat covariates and *in situ* diver or video survey data of Uku. Output from the SDMs were used to create maps delineating Uku EFH including continuous probability of occurrence maps as well as EFH categorical maps. The resulting categorical maps allow the geographic identification of “hot spot”, “core habitat”, and “basic EFH” areas for Uku EFH throughout the main Hawaiian Islands. These analyses represent the first model-based approach to delineating EFH for a U.S. Western Pacific stock and could serve as a framework for the EFH descriptions of other managed species in the region.

MATERIAL AND METHODS

Study Area

The study area included the “shallow” (0 to 30 m depth) and “deep” (30 m to 300 m) seafloor around the 8 main Hawaiian Islands (MHI). The Hawaiian archipelago encompasses a group of volcanic islands and atolls that span 2500 km in the central north Pacific Ocean (Figure 2). The geography of these volcanic islands is characterized by prominent coastal capes and headlands that demarcate coastal exposures to different climate and ocean conditions. The north coasts of Kauai, Oahu, and Maui are exposed to large northern hemisphere winter swells (≥7 m), while southern hemisphere storms produce waves (3 to 5 m) along Hawaiian south shores in summer (Fletcher et al. 2008). The eastern or windward side of the islands experience consistent easterly trade winds (10 to 20 knots) that generate steady wind-driven waves (1 m; Fletcher et al. 2008). There are only 2 large, natural semi-enclosed waters bodies in the MHI, Pearl Harbor and Kaneohe Bay on Oahu (Fig. 1). In shallow waters (0 – 30 m), coral reefs are found around the coasts and embayments of all islands (Battista et al. 2007). In deeper waters (30 – 300 m), a mix of softbottom and hardbottom habitats are found through the region with coral reefs are less common at these depths.

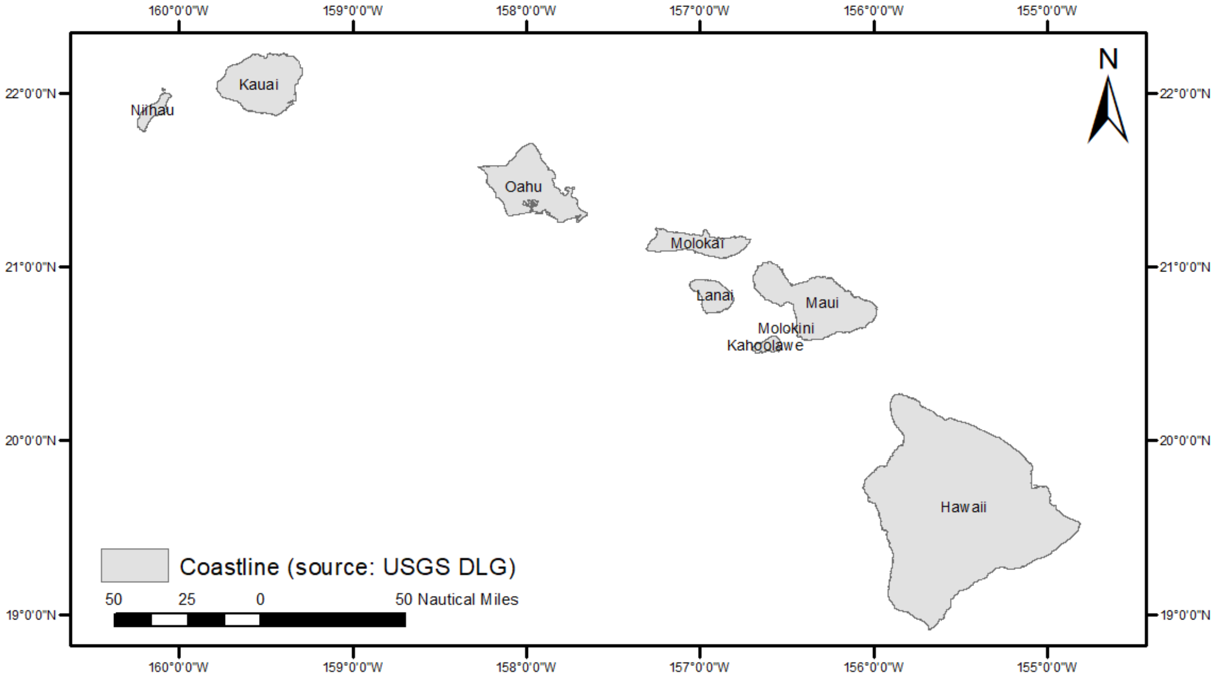


Figure 2. A map of the eight Main Hawaiian Islands.

Literature and Data Search

I searched the scientific literature (i.e., Google Scholar and Web of Science) and reviewed fishery-independent survey data information relevant to the delineation of EFH for *Aprion virescens* in the main Hawaiian Islands. Here, the main Hawaiian Islands are defined as the area of the US EEZ that encompasses the Niihau, Kauai, Oahu, Maui Nui, and Hawaii. GIS bathymetry and habitat layers were provided by the Pacific Islands Benthic Habitat Mapping Center (https://www.soest.hawaii.edu/pibhmc/cms/), the Hawaii Mapping Research Group (http://www.soest.hawaii.edu/HMRG/cms/about-hmrg/), and the NOAA Centers for Coastal Ocean Science (https://coastalscience.noaa.gov/research/project-explorer/). The NOAA Pacific Islands Fishery Science Center (PIFSC), The Nature Conservancy, and the Hawaii State Division of Aquatic Resources provided diver survey data for Uku. NOAA PIFSC and Jake Asher provided baited stereo-video camera survey data for Uku.

Survey Data Sources

A number of survey data sources were reviewed for Uku observations in the main Hawaiian Islands between 2010 and 2019. Two primary field methods were used for the direct observation of Uku in fishery-independent surveys, diver surveys (Heenan et al. 2017) and baited stereo-video cameras (Merritt et al. 2011, Amin et al. 2017, Asher et al. 2017). The details of these data sources are provided in the following sections.

Stationary Diver Fish Surveys - NOAA Fisheries Pacific Island Fisheries Science Center (PIFSC)

- Number of surveys: 1,682
- Surveys with *A. virescens* observed: 252 (14.9% of total)
- Year range: 2010-2019
- Island areas surveyed (8): Niihau, Kauai, Oahu, Molokai, Lanai, Maui, Kahoolawe, Hawaii
- Depth range: 1.3 m – 30.0 m
- Median uku length (and range): 54 cm (22 cm – 110 cm) FL
- Methods: Visual observations of fish community by SCUBA divers using the stationary point count method in shallow coral reef habitats (Heenan et al. 2017). Each survey represents data averaged from multiple diver surveys, typically two per site. Uku observations were collected along with observations of multiple species of fish. The number and length of uku were visually estimated by divers, not measured directly.
- Comments: These surveys were performed during 2010, 2012, 2013, 2015, 2016, and 2019. Not all islands and coastlines were surveyed in each year (Figure 3).
- Contact: Tye Kindinger, tye.kindinger@noaa.gov

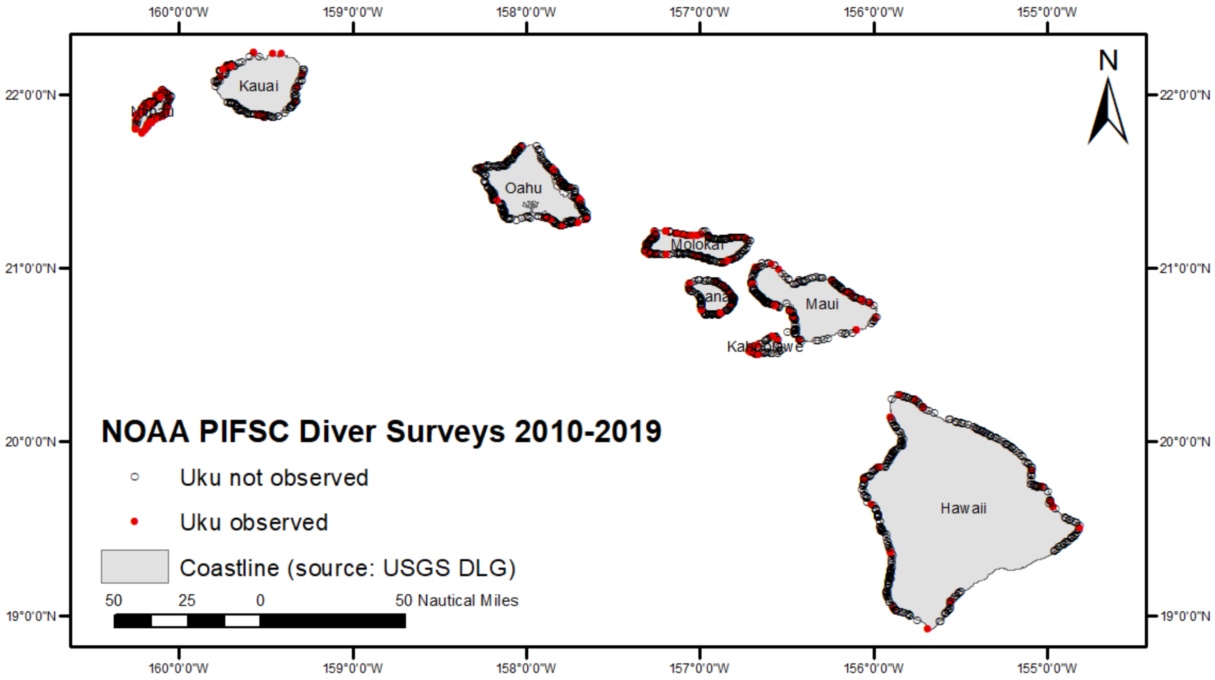


Figure 3. NOAA PIFSC diver surveys for uku (*Aprion virescens*) in the main Hawaiian Islands, 2010-2019. Surveys include those with uku observed (red circles) and those with no uku observed (open circles).

Diver Transect Surveys (FAHU) – Hawaii Division of Aquatic Resources (DAR)

- Number of surveys: 2,395
- Surveys with *A. virescens* observed: 127 (5.3% of total)
- Year range: 2015-2020
- Island areas surveyed (3): Kauai, Lanai, Maui
- Depth range: 0.6 m – 22.9 m
- Median uku length (and range): 40 cm (12.5 cm – 120 cm) FL
- Methods: Visual observations of fish community by SCUBA divers using the belt transect method in shallow coral reef habitats (cite?). The method used 25 m x 5 m belt transects. Uku observations were collected along with observations of a multiple coral reef species. The number and length of uku were visually estimated by divers, not measured directly. Results from survey replicates were averaged together for each site.
- Comments: These surveys were performed annually during 2015-2020. Not all coastlines were surveyed in each year (Figure 4).
- Contact: Laura Gajdzik, lgajdzik@hawaii.edu

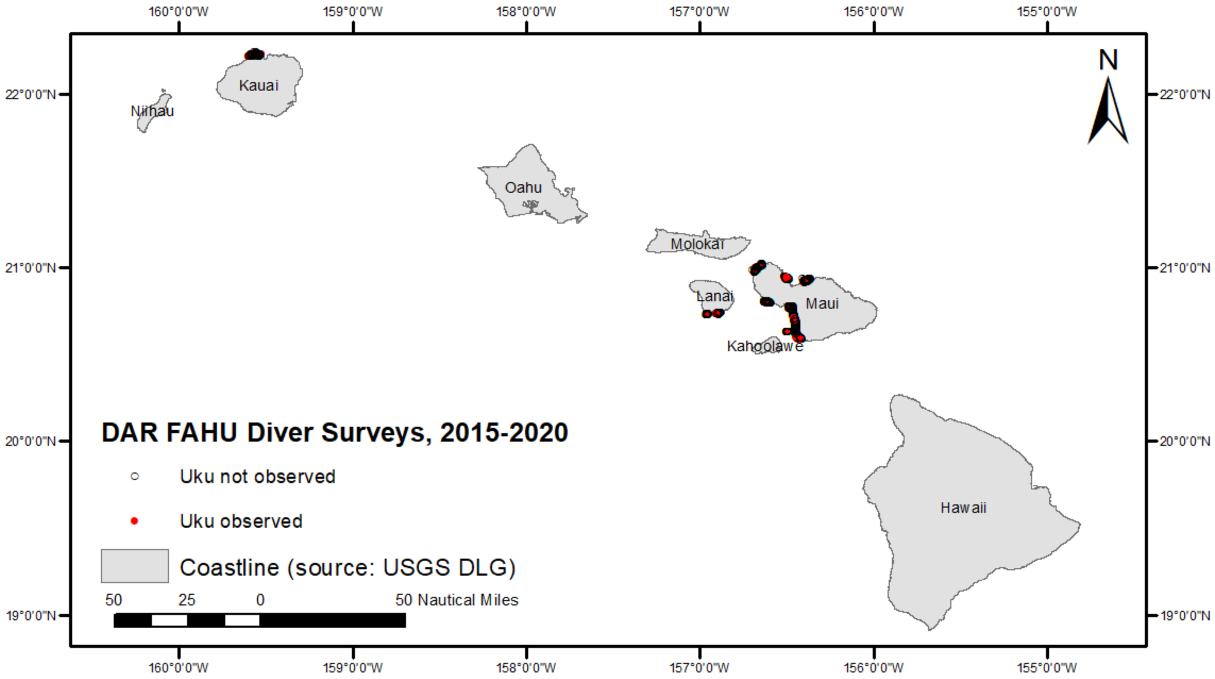


Figure 4. DAR FAHU diver surveys for uku (*Aprion virescens*) in the main Hawaiian Islands, 2015-2020. Surveys include those with uku observed (red circles) and those with no uku observed (open circles).

Diver Transect Surveys – The Nature Conservancy

- Number of surveys: 2,730
- Surveys with *A. virescens* observed: 112 (4.1% of total)
- Year range: 2012-2020
- Island areas surveyed (4): Oahu, Maui, Kahoolawe, Hawaii
- Depth range: 0 m – 60.6 m
- Median uku length (and range): 40 cm (6.3 cm – 130 cm) FL
- Methods: Visual observations of fish community by SCUBA divers using the belt transect method in shallow coral reef habitats. The method used 25 m x 5 m belt transects. Uku observations were collected along with observations of a multiple coral reef species. The number and length of uku were visually estimated by divers, not measured directly. Results from survey replicates were averaged together for each site.
- Comments: These surveys were performed during 2012-2020 but not necessarily every year at all locations (Figure 5).
- Contact: Eric Conklin, econklin@tnc.org

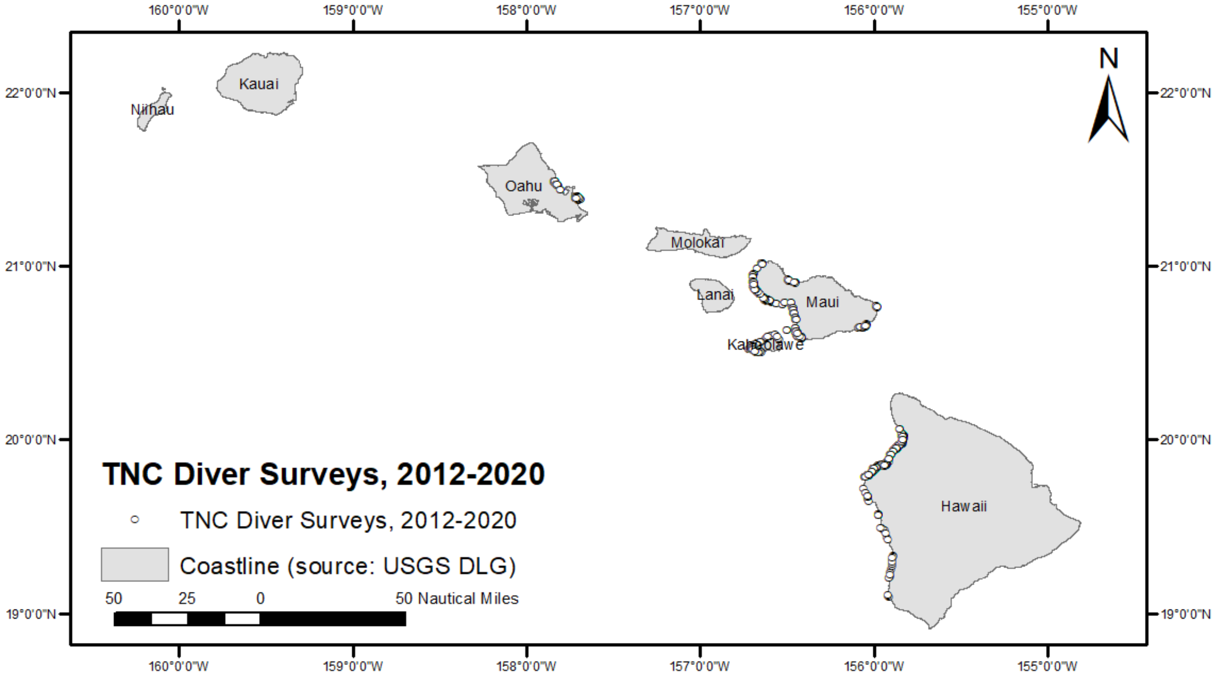


Figure 5. The Nature Conservancy diver surveys for uku (*Aprion virescens*) in the main Hawaiian Islands, 2012-2020. Due to data confidentiality, survey locations with uku observations cannot be identified.

Towed Diver Fish Surveys - NOAA Fisheries Pacific Island Fisheries Science Center

- Number of surveys: 1,722
- Surveys with *A. virescens* observed: 44 (2.6% of total)
- Year range: 2010-2016
- Island areas surveyed (7): Niihau, Kauai, Oahu, Molokai, Lanai, Maui, Hawaii
- Depth range: 10 m – 20 m
- Median uku length range: 62.5 cm (50 cm – 90 cm) FL
- Methods: Visual observations of fish community by SCUBA divers using the stationary point count method in shallow coral reef habitats (cite).
- Comments: These surveys were performed during 2010 and 2016 (Figure 6). Each survey represents data collected during a tow “segment”. The centroid of the tow segment was used to identify the geographic coordinates for the survey location. Uku observations were collected along with observations of a multiple coral reef species. The number and length of uku were visually estimated by divers, not measured directly.
- Contact: Tye Kindinger, tye.kindinger@noaa.gov

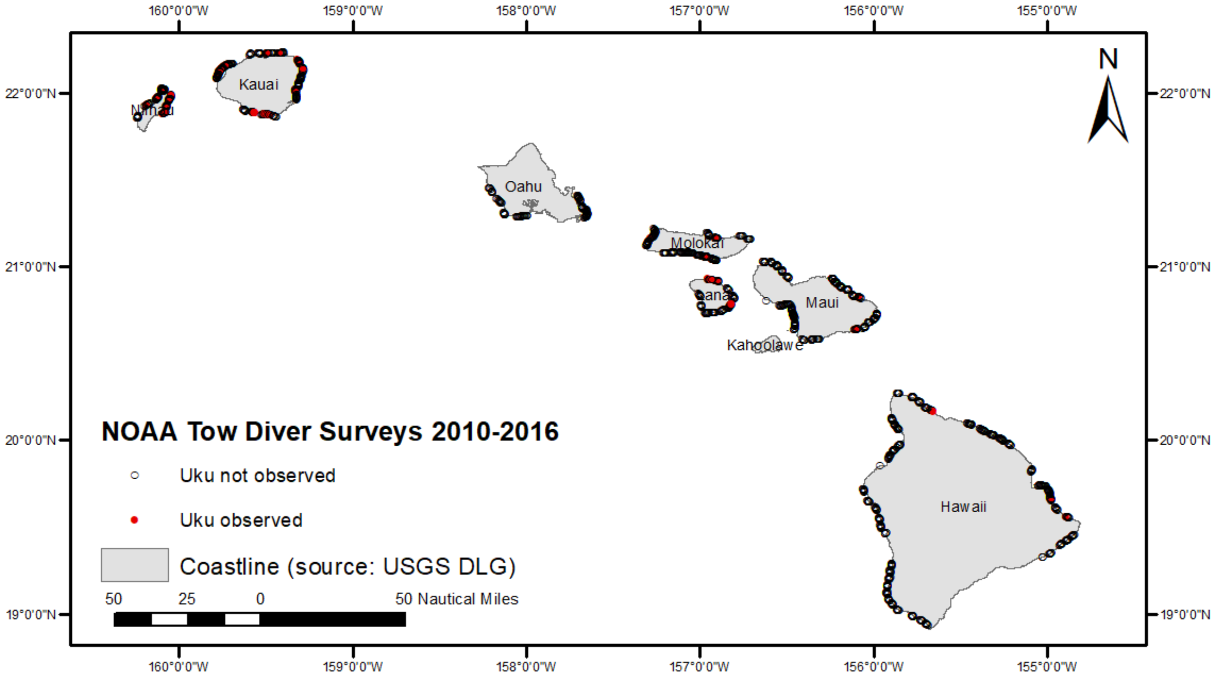


Figure 6. NOAA PIFSC towed diver surveys for uku (*Aprion virescens*) in the main Hawaiian Islands, 2010-2016. Surveys include those with uku observed (red circles) and those with no uku observed (open circles).

Baited Remote Underwater Video (BRUVs) Surveys

- Contact: Jacob Asher, jakeasher@hotmail.com
- Number of surveys: 107
- Surveys with *A. virescens* observed: 28 (26.2% of total)
- Year range: 2010-2013
- Island areas surveyed (4): Oahu, Molokai, Lanai, Maui
- Depth range: 2.7 m – 96.6 m
- Uku length range: 22.2 cm – 107.2 cm FL
- Methods: Baited remote underwater video (BRUVs) surveys recorded visual observations of the fish community using stationary stereo-video baited camera arrays in shallow to mesophotic hardbottom and softbottom habitats (Asher et al. 2017). Uku observations were collected along with observations of multiple fish species. The maximum number and length of uku were visually estimated from the videos using software tools (i.e., Event Measure), not measured directly.
- Comments: These surveys were performed during 2010-2013 (Figure 7).

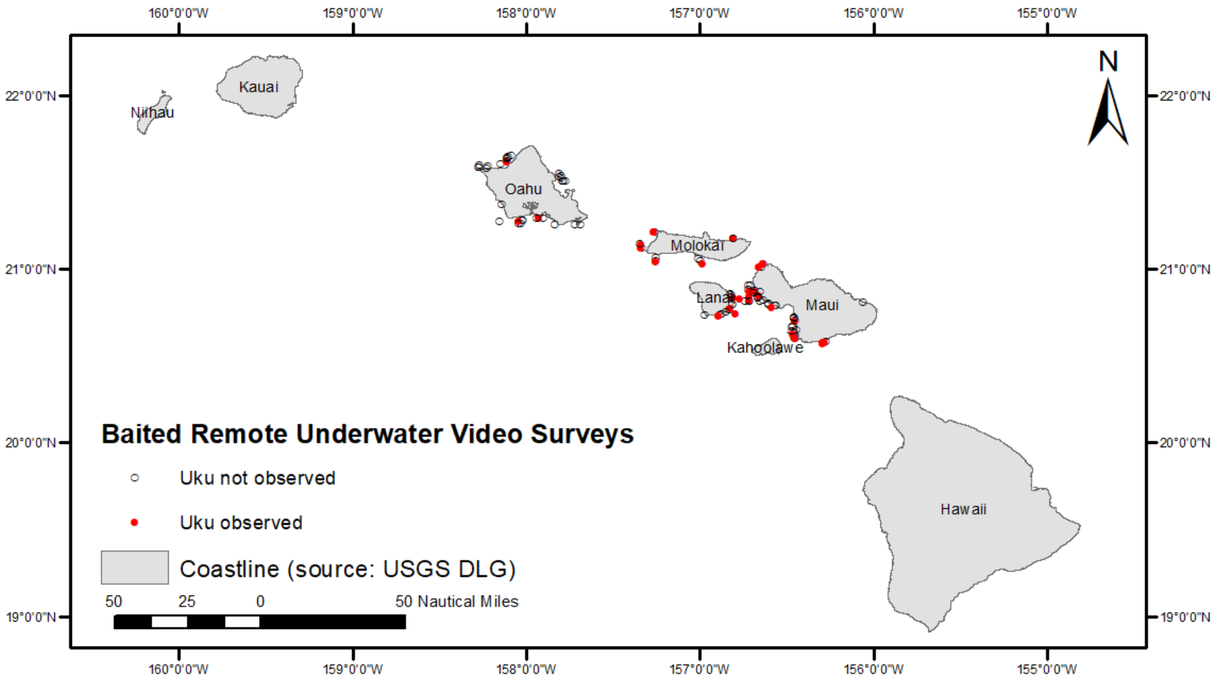


Figure 7. Baited remote underwater videos (BRUVs) for uku (*Aprion virescens*) in the main Hawaiian Islands, 2010-2013. Surveys include those with uku observed (red circles) and those with no uku observed (open circles).

BotCAM Video Surveys – NOAA PIFSC / University of Hawaii

- Number of surveys: 465
- Surveys with *A. virescens* observed: 24 (5.2% of total)
- Year range: 2011-2014
- Island areas surveyed (3): Lanai, Maui, Kahoolawe
- Median depth (and range): 164.5 m (63.7 m – 314.0 m)
- Uku length range: 52.7 cm – 75.3 cm FL
- Methods: Baited underwater video surveys recorded visual observations of the fish community using stationary stereo-video baited camera arrays in mesophotic hardbottom and softbottom habitats (Merritt et al. 2011).
- Comments: These surveys were performed annually during 2011-2014 (Figure 8). Uku observations were collected along with observations of multiple fish species. Surveys with "at least 1" uku were given a value of 1 for Nmax. Survey records with a *Species* identification of "Lutjanid/ae", "Perciformes", "Teleost", or "too dark to annotate" were not included in the analysis.
- Contact: Audrey Rollo, audrey.rollo@noaa.gov

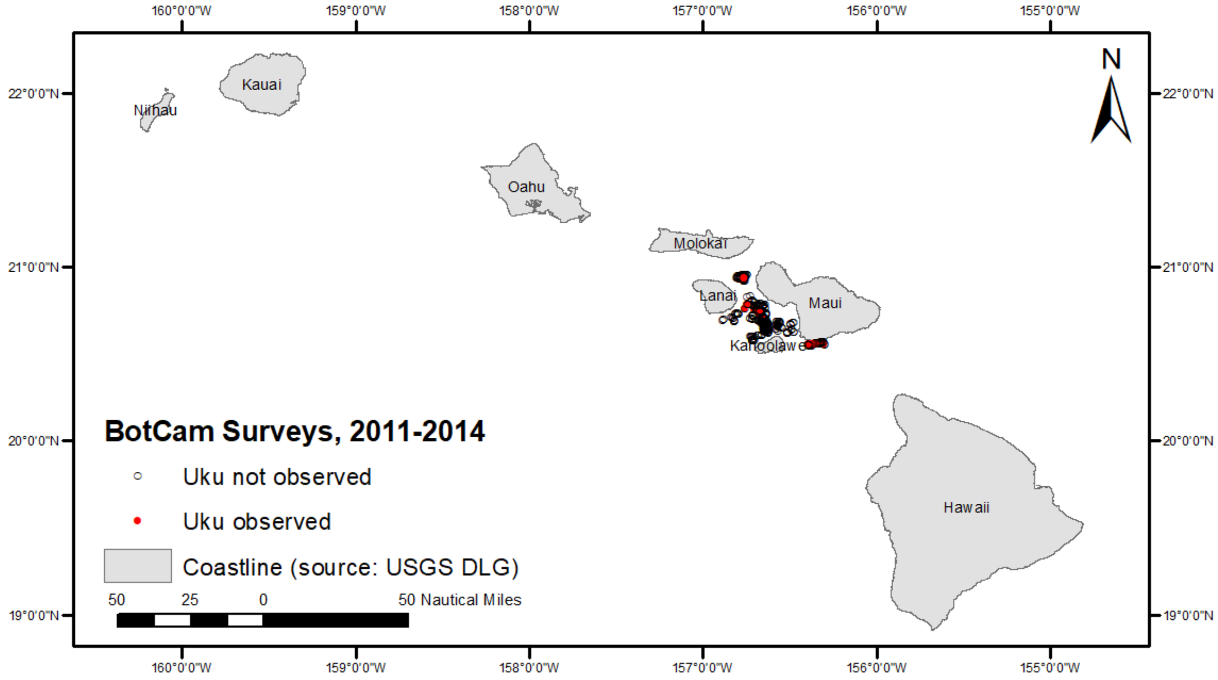


Figure 8. BotCam underwater video surveys for uku (*Aprion virescens*) in the main Hawaiian Islands, 2011-2014. Surveys include those with uku observed (red circles) and those with no uku observed (open circles).

MOUSS Video Surveys – NOAA PIFSC

- Number of surveys: 951
- Surveys with *A. virescens* observed: 105 (11.0% of total)
- Year range: 2016-2019
- Island areas surveyed (8): Niihau, Kauai, Oahu, Molokai, Lanai, Maui, Kahoolawe, Hawaii
- Median depth (and range): 159.5 m (44.8 m – 291.7 m)
- Uku length range: NA
- Methods: Baited underwater video surveys recorded visual observations of the fish community using stationary stereo-video camera arrays in shallow to mesophotic hardbottom and softbottom habitats (Amin et al. 2017).
- Comments: These surveys were performed annually during 2016-2019 (Figure 9). Uku observations were collected along with observations of multiple fish species. Surveys with "at least 1" uku were given a value of 1 for Nmax. Survey records with a *Species* identification of "Lutjanid/ae", "Perciformes", "Teleost", or "too dark to annotate" were not included in the analysis.
- Contact: Audrey Rollo, audrey.rollo@noaa.gov

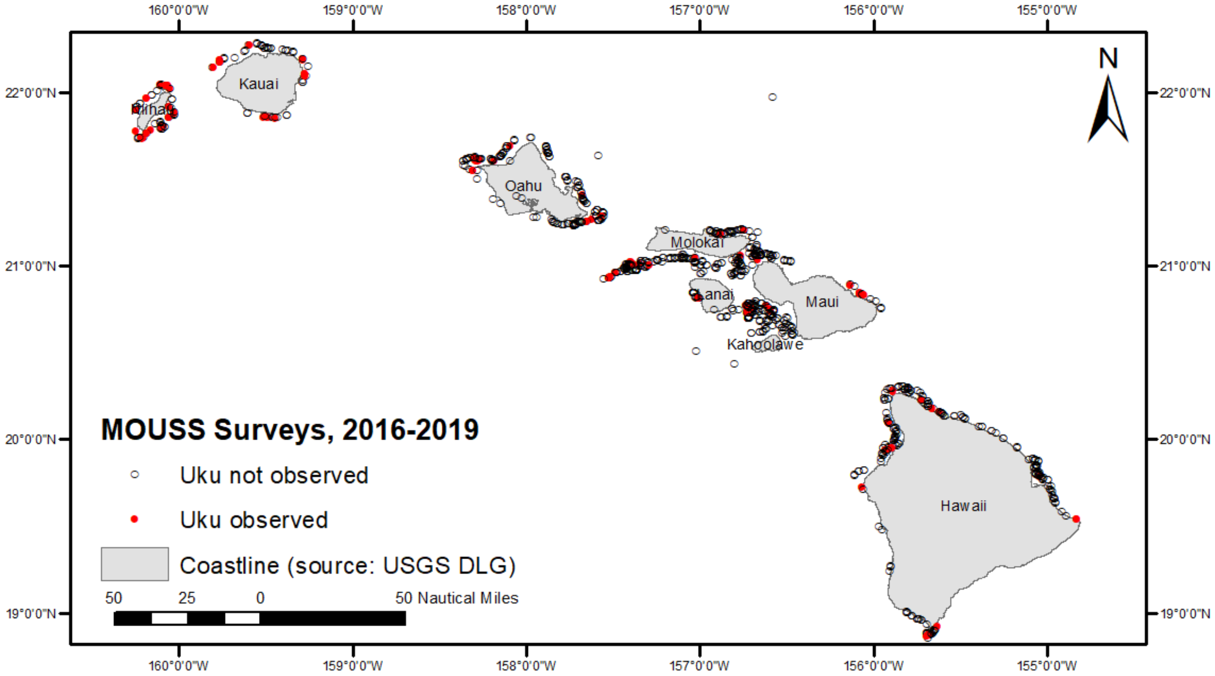


Figure 9. MOUSS underwater video surveys for uku (*Aprion virescens*) in the main Hawaiian Islands, 2011-2014. Surveys include those with uku observed (red circles) and those with no uku observed (open circles).

Survey Data Used for SDM Models

While several potential survey data sources were evaluated for inclusion in the SDMs (see prior sections), only the NOAA PIFSC SPC diver surveys for shallow habitats and NOAA/UH baited remote underwater stereo-video surveys (i.e., BotCam, MOUSS, & BRUVs) for deep habitats were used. These surveys were selected based on their methodological consistency, well-documented field techniques, and data collection QA/QC procedures by NOAA scientists. Due to differences in survey data collection methods between diver surveys and stereo-video camera surveys, separate SDM models for Uku were constructed for shallow (0-30 m depth) depths using NOAA fish SPC diver surveys and deep (30-300 m) depths using NOAA and UH baited stereo-video camera arrays. Shallow diver surveys included information on fish lengths but most of the deep video surveys did not include length information for observed Uku. The lack of length data for most of the surveys meant that sufficient information was not available to create separate EFH maps for sub-adult and adult fish. The EFH analyses and maps that follow are for aggregated sub-adult/adult life stages. Some stereo-video camera surveys were performed in depths shallower than 30 m but these surveys were not included in the SDMs. A few surveys were located on land or in water deeper than 300 m. These were not included in the models and required additional QA/QC by the primary data providers. Shallow and deep survey sites were present around Niihau and Kauai (Figure 10), Oahu (Figure 11), Maui Nui (Figure 12), and Hawaii (Figure 13).

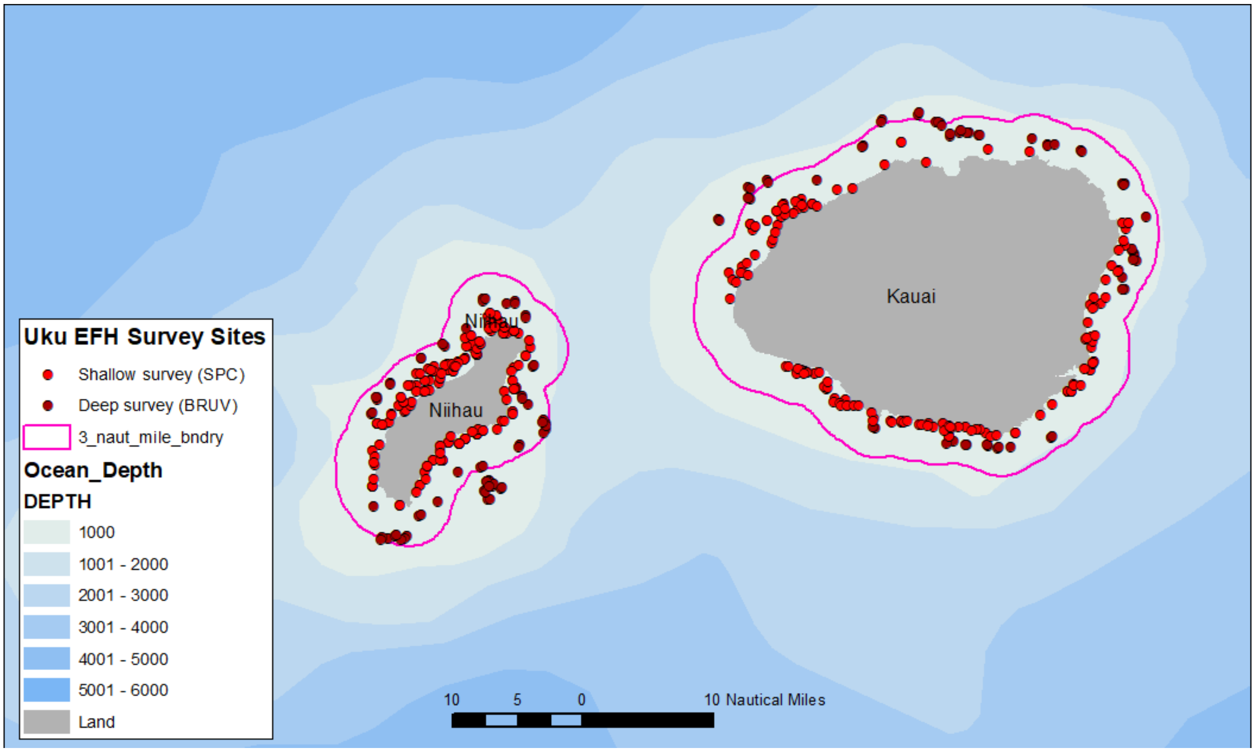


Figure 10. Shallow and deep survey sites for Uku (*Aprion virescens*) in Niihau and Kauai.

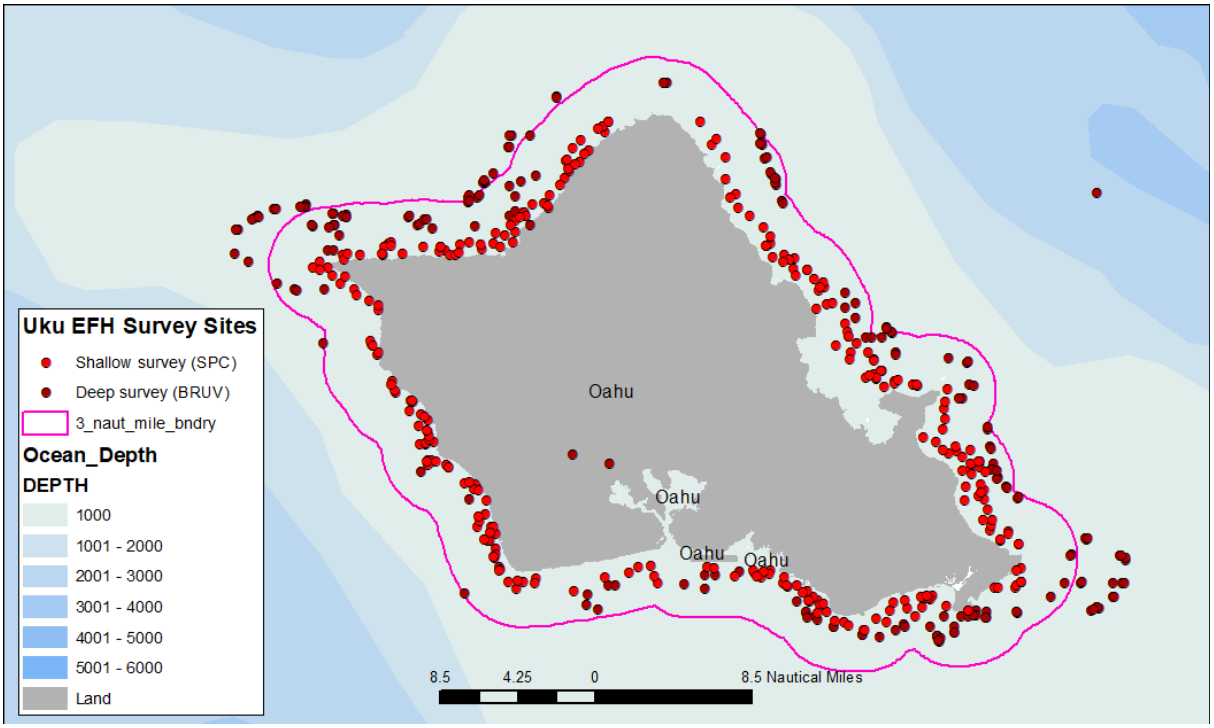


Figure 11. Shallow and deep survey sites for Uku (*Aprion virescens*) around Oahu.

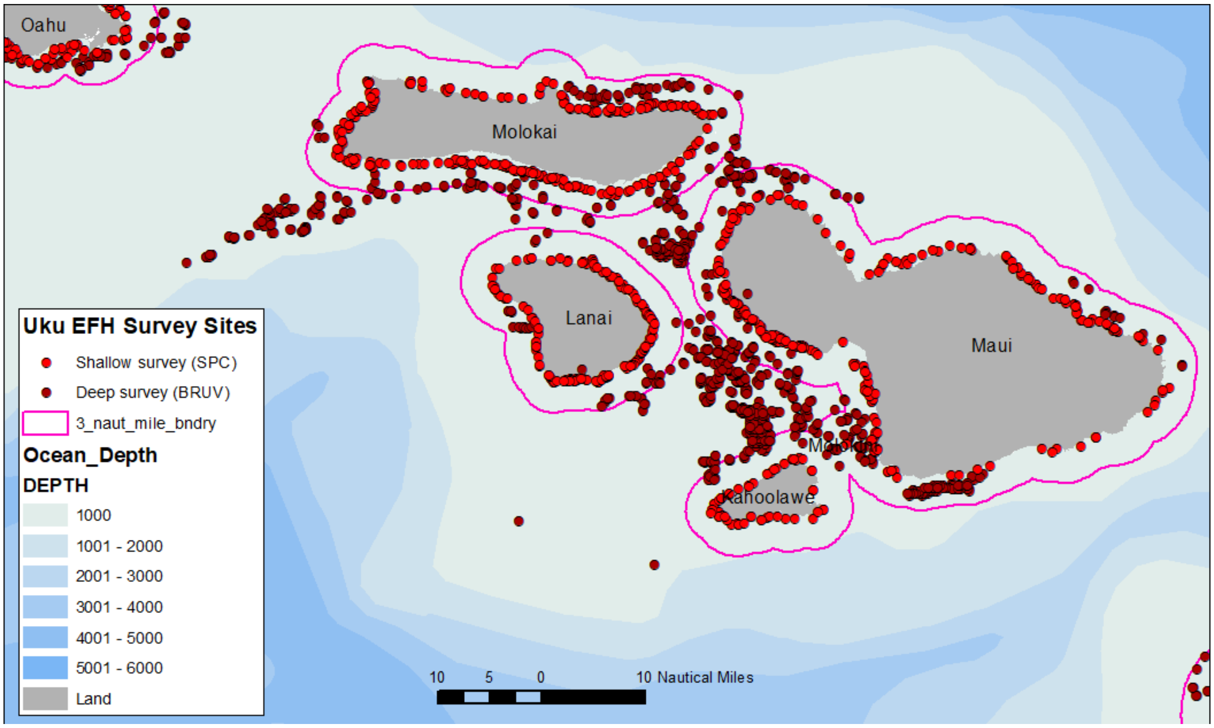


Figure 12. Shallow and deep survey sites for Uku (*Aprion virescens*) around Maui Nui.

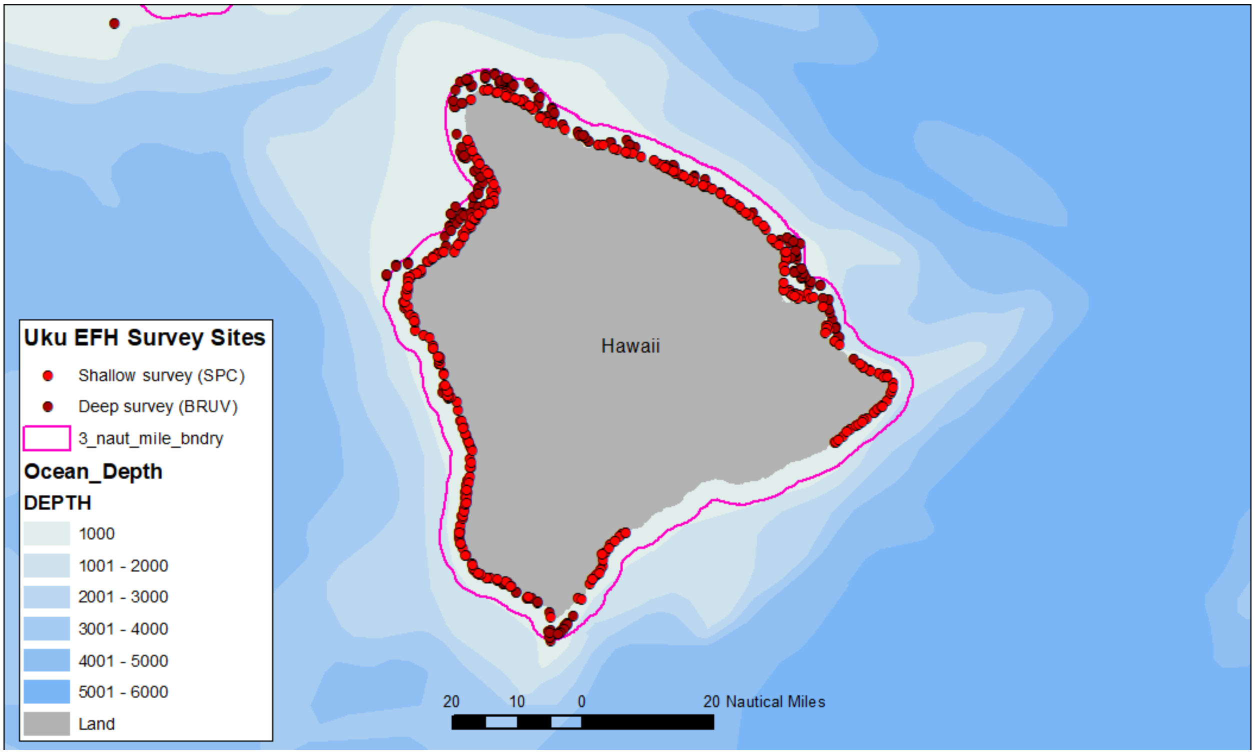


Figure 13. Shallow and deep survey sites for Uku (*Aprion virescens*) around Hawaii Island.

Habitat Covariates

A set of habitat covariates were used to parameterize and select the best fit SDMs. The covariates represented observed, derived and modeled variables that were expected to influence the distribution of Uku. For all covariates, the values were resampled to regular spatial raster grids of 50 m x 50 m resolution. Bathymetry-derived variables—slope, aspect, and rugosity—were calculated in ArcGIS (V.10.7.1) with an eight-cell neighborhood (Burrough and McDonnell, 1998). Terrain ruggedness, referred to as rugosity hereafter, was calculated with an eight-cell neighborhood using ArcGIS Benthic Terrain Modeler (Wright et al., 2005) and ranged from 0 (no variation) to 1 (complete variation). Bathymetry-derived variables were calculated at 50-m resolution of bathymetry provided by the Hawaii Mapping Research Group's bathymetry synthesis. As an example of a "dynamic" habitat variable, for the shallow model only, SWAN wave model output of maximum significant wave height and mean significant wave height were estimated. Collinearity among covariates was examined prior to using them in the SDMs. Paired correlations for all covariates were below $r = 0.6$, an acceptable threshold for boosted regression tree models (Elith et al. 2008).

Species distribution modeling

SDM modeling adapted methods used by Franklin et al. (2013) and Oyafuso et al. (2017). Boosted regression tree (BRT) models were constructed for Uku occurrence (i.e., presence/absence) using the routines *gbm* (generalized boosted regression models) v2.1.8 (Ridgeway 2020) and *gbm.step* (Elith et al. 2008) in the R statistical program V4.03 (R

Development Core Team, www.r-project. org). BRT models combine re gression trees that fit environmental predictors to response variables with a boosting algorithm that assembles an ensemble of trees in an additive, stage-wise fashion (Hastie et al. 2001, Elith et al. 2008). Within the BRT models, 3 terms were used to optimize predictive performance: tree complexity, learning rate, and bag-fraction. Tree complexity *(tc)* determined the number of nodes in a tree that should reflect the true interaction order on the response being modeled, although this is often unknown, and learning rate *(lr)* was used to shrink the contribution of each tree as it is added to the model (Elith et al. 2008). The bag-fraction determined the proportion of data to be selected at each step and, therefore, the model stochasticity (Elith et al. 2008). For each species, the BRT model training dataset was a one-time random selection of 70% of the original total dataset of model grid cells (Table 1). The remaining 30% was held out for independent validation of each optimal BRT model. I determined optimal settings for these parameters by examining the cross-validation deviance over *tc* values 1−5, *lr* values of 0.01, 0.05 and 0.001, and bag fractions of 0.5 and 0.75. All possible combinations were run, with the optimal number of trees in each case being determined by *gbm.step* (Elith et al. 2008). Each model run included 10-fold cross-validation using training data sets. The combination of the 3 parameter settings with the lowest cross-validation deviance was then selected to produce the optimal BRT model for each species fit with the entire training dataset (Elith et al. 2008). Finally, the deviance of the optimal model was evaluated on the test (30%) dataset. All models were run with binary measures of Uku presence (i.e., 0 or 1) which were treated as a binomial response distribution. For the final BRT models, the relative contribution of each predictor was based on the number of times the variable was selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman & Meulman 2003, Elith et al. 2008). Partial dependency plots were used for interpretation and to quantify the relationship between each predictor variable and response variable, after accounting for the average effect of all other predictor variables in the model. I used *gbm.interactions* (Elith et al. 2008) to quantify interaction effects between predictors. The relative strength of interaction fitted by BRT was quantified by the residual variance from a linear model, and the value indicates the relative relative degree of departure from a purely additive effect, with zero indicating no interaction effects fitted (Elith et al. 2008). I defined a threshold interaction value and reported the interactions with values ≥ 0.1.

SDM Evaluation

A set of common evaluation metrics of predictive performance was calculated on the models fitted to the test datasets. Area Under the Receiving Operating Curve (AUC) calculates the ability of a model to discriminate between a presence or absence observation. Values of AUC are coarsely interpreted as: bad: 0.50–0.59; poor: 0.60–0.69; fair: 0.70–0.79; good: 0.80–0.89; excellent: 0.90–1.0 (Hosmer et al., 2013). Specificity, Sensitivity and the True Skill Statistic (TSS) were calculated using a probability threshold that balances sensitivity and specificity similar to Schroeder and Richter (2000). True Skill Statistic values range from −1 to +1 where values <0 indicating a predictive model worse than random, zero indicating an indiscriminate predictive model, and +1 indicating a perfect predictive model. Lastly, percent deviance explained was calculated as: %*Deviance Explained* = 100%(1 − *Residual Deviance*/*Null Deviance*).

RESULTS

Uku was a habitat generalist found in both hardbottom and softbottom habitats with most observations occurring in depths between 30-120 m. The species seems to be a solitary predator, with most positive observations of only single fish (e.g., only 18 of 156, 12%, of deep camera observations had more than one Uku). The final optimal BRT models for shallow and deep habitats predicted the habitat suitability of combined sub-adult/adult Uku probability of occurrence for the MHI. Model parameter settings for the optimal BRT models were selected from a set of possible combinations of bag fraction, tree complexity, and learning rate (Table 1). The optimal models for shallow and deep habitats had “good” model fits based on AUC and the Total Skill Statistic with acceptable performances in model specificity and sensitivity (Table 2). The occurrence of Uku was higher on N-NW and SE-SW facing habitats and increased with decreasing depths and increasing wave heights in the shallow model (0-30 m) (Figure 14). Ranked relative importance of variables for the optimal shallow model was aspect (25.2% of relative importance), depth (24.4%), maximum wave height (16.8%), mean wave height (12.3%), slope (9.7%), rugosity (8.4%), and sand (3.2%) (Figure 14). For the deeper model (30-300 m), the occurrence of Uku was higher in depths shallower than 100 m and hard-bottom habitats (Figure 15). Ranked relative importance of variables for the optimal deep model was depth (44.9%), sand (17.2%), slope (13.1%), aspect (12.7%), and rugosity (12.2%).

Table 1. Parameter values selected for shallow and deep optimal BRT models. Abbreviations are for bag fraction (bf), tree complexity (tc), learning rate (lr), number of trees (trees), and mean deviance for cross validated model runs (devmean)

Model	bf	tc	lr	trees	devmean
Shallow	0.5	4	0.001	4400	0.537
Deep	0.5	4	0.001	3300	0.488

Table 2. Model performance values for the shallow and deep optimal BRT models. Abbreviations are for proportion of deviance explained (Environ_DEV), area under curve-receiver operating characteristic (AUC), total skill statistic (TSS), and model specificity and sensitivity.

Model	Environ_DEV	AUC	TSS	Specificity	Sensitivity
Shallow	0.42	0.82	0.43	0.83	0.61
Deep	0.39	0.86	0.55	0.76	0.79

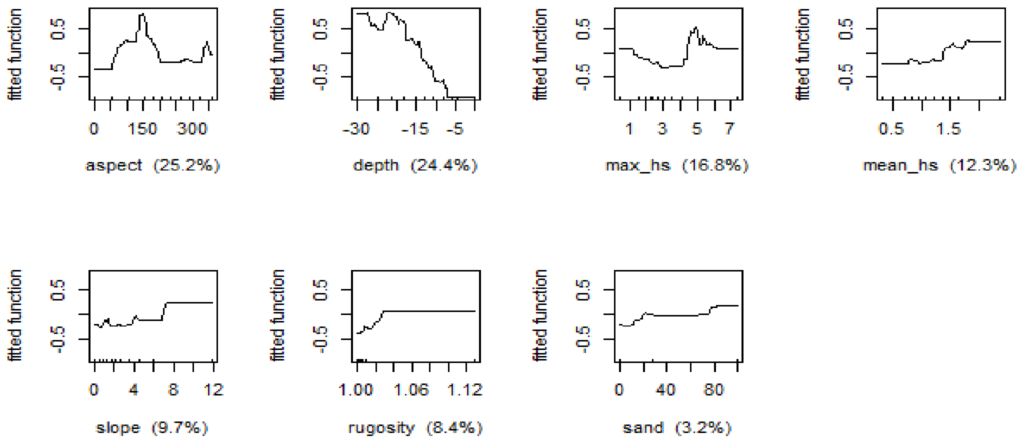


Figure 14. Partial regression plots for covariates in the optimal “shallow” BRT model.

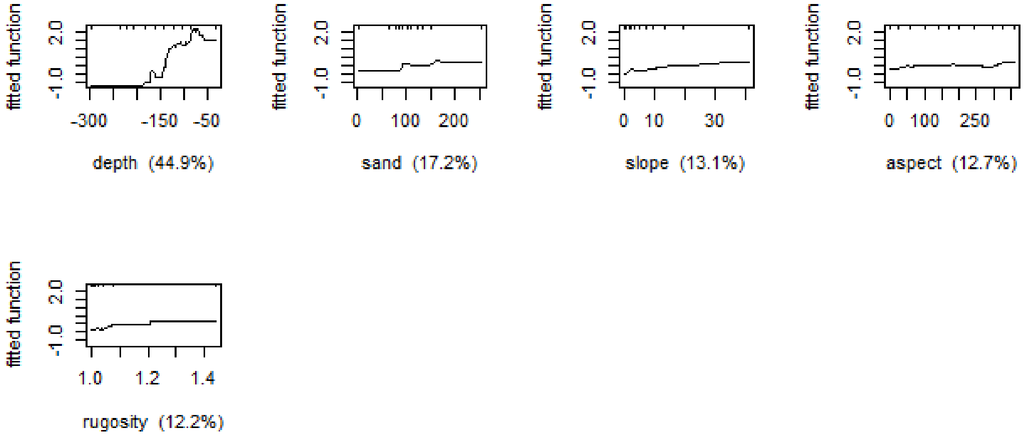


Figure 15. Partial regression plots for covariates in the optimal “deep” BRT model.

Uku EFH Habitat Suitability Maps

The optimal shallow (0-30 m) and deep (30-300 m) BRT models were used to predict Uku sub-adult/adult occurrence to the waters around the main Hawaiian Islands. The GIS maps were generated by using the optimal models fit to the values of habitat covariates from each geographic location in the main Hawaiian Islands domain (in R) to generate a predicted probability of Uku occurrence. Across the shallow habitat range, the mean probability of occurrence per habitat cell was 0.09 (range = 0.02 – 0.71, sd = 0.09, n = 477,795 habitat cells). For the deeper habitat range, the mean probability of occurrence per habitat cell was 0.12 (range = 0.01 – 0.88, sd = 0.12, n = 2,259,733 habitat cells). The maps show the model-predicted probability of occurrence for sub-adult/adult Uku for Kauai and Niihau (Figure 16), Oahu (Figure 17), Maui Nui (Figure 18), and Hawaii island (Figure 19).

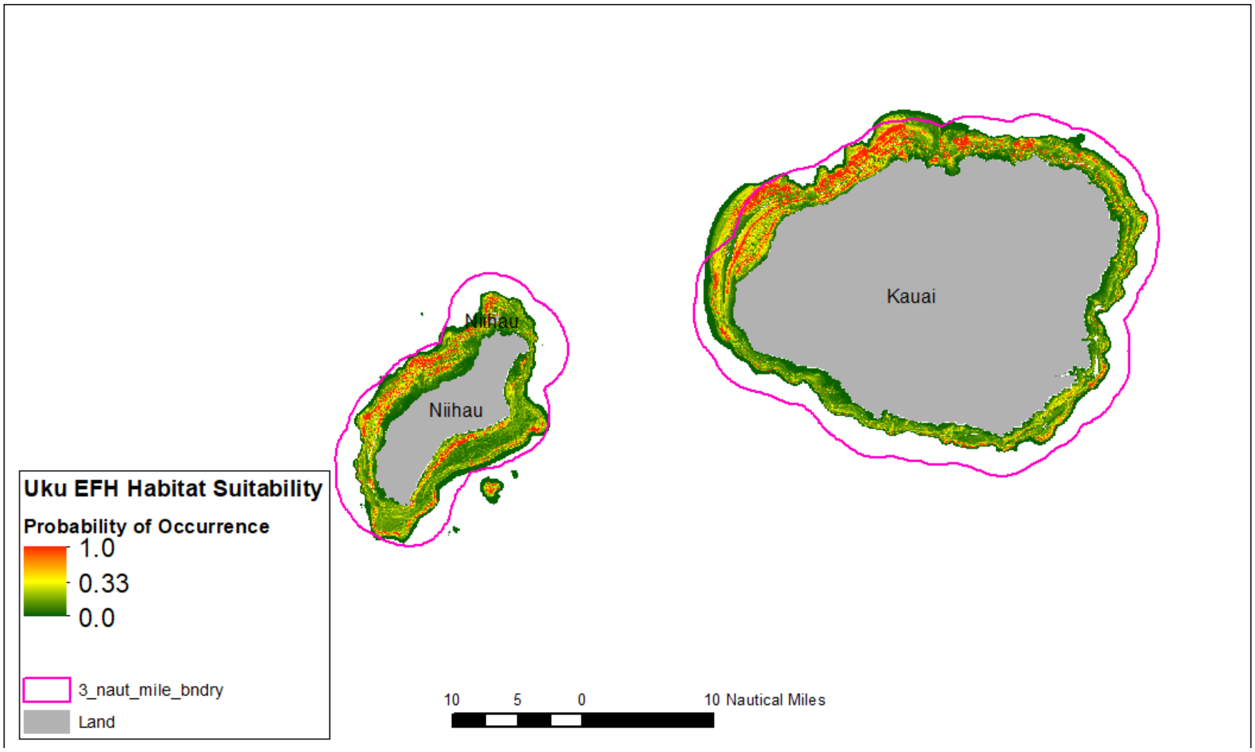


Figure 16. Uku (*Aprion virescens*) predicted probability of occurrence around Kauai and Niihau from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

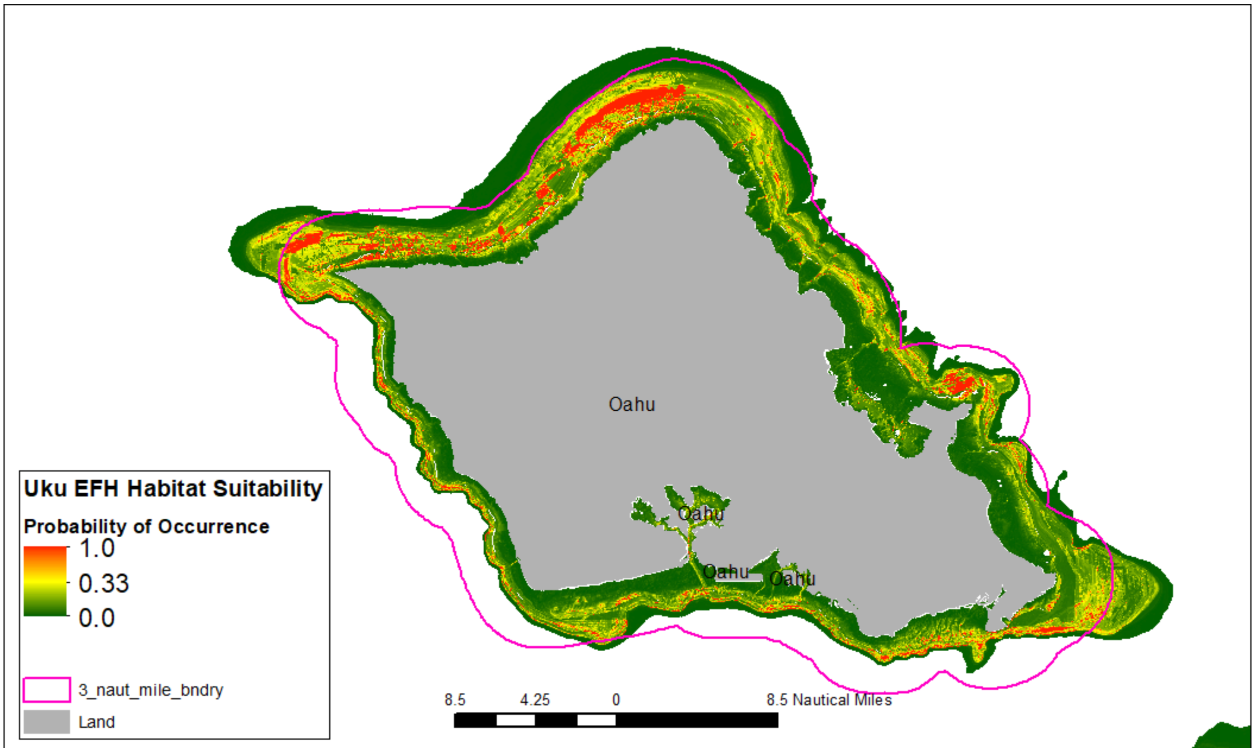


Figure 17. Uku (*Aprion virescens*) predicted probability of occurrence around Oahu from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

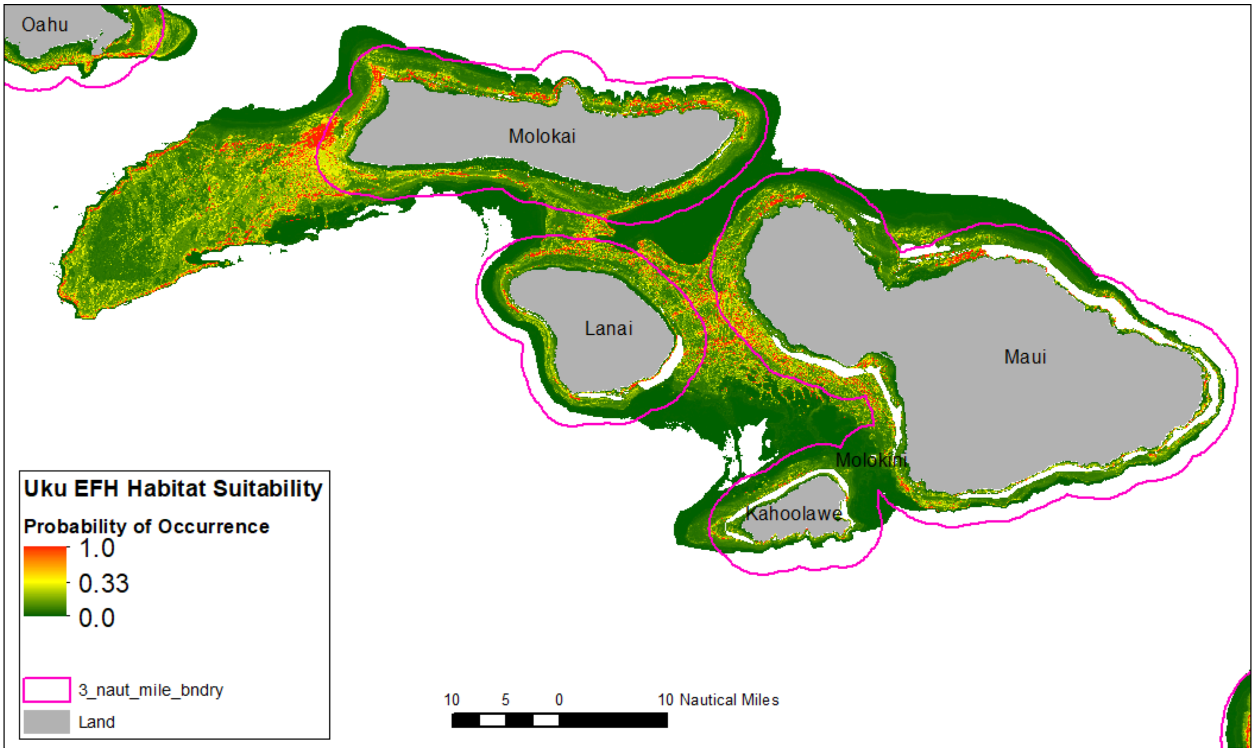


Figure 18. Uku (*Aprion virescens*) predicted probability of occurrence around Maui Nui from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

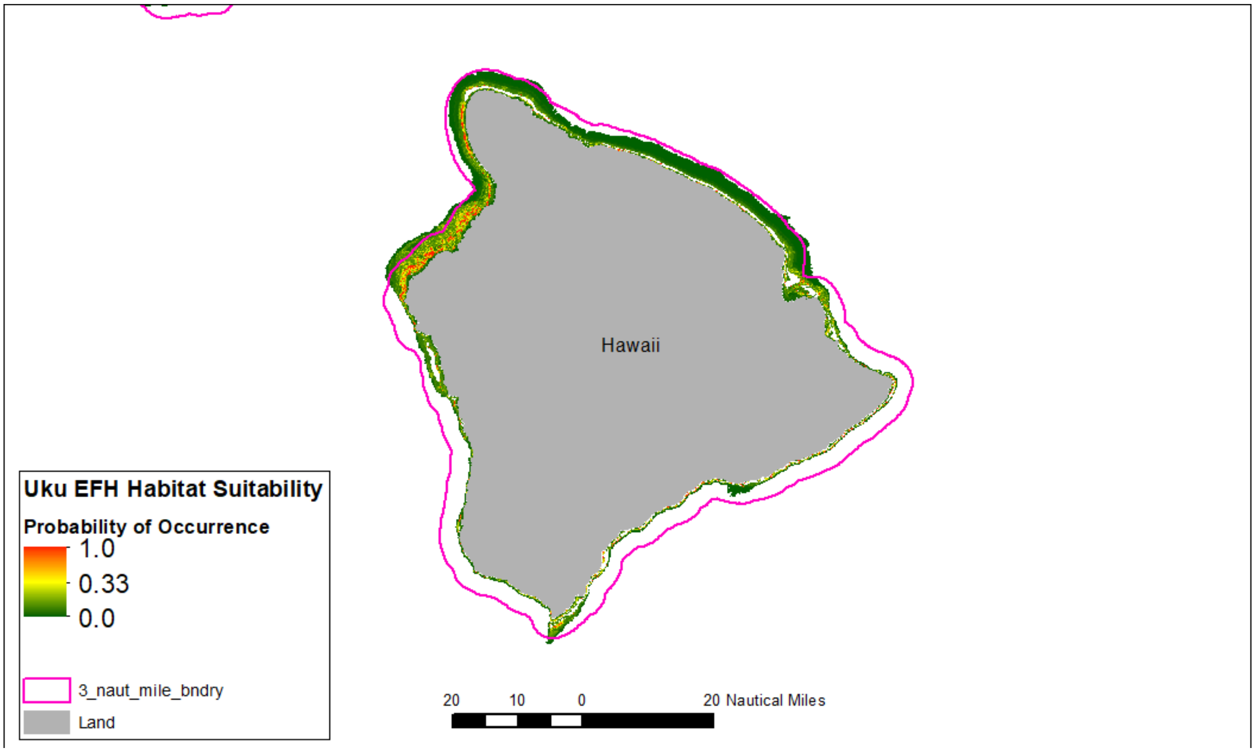


Figure 19. Uku (*Aprion virescens*) predicted probability of occurrence around Hawaii island from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

Uku EFH Categorical Maps

Maps of the predicted probability of occurrence from habitat suitability models display a continuous scale of the response variable that is difficult to interpret for decision making. To better visualize the relative importance of different habitat areas, a categorical method for EFH using the absolute 25%, 50% and 95% quantiles of the Uku probability of occurrence is introduced. This approach has been utilized for delineation of EFH in other US Fishery Management Regions such as the North Pacific (Pirtle et al. 2020). The quantiles identified “hot spots” (predicted probability of occurrence: 1.0 – 0.75), “core habitat” (0.75 – 0.50), and “basic EFH” (0.50 – 0.05) for waters around Kauai and Niihau (Figure 20), Oahu (Figure 21), Maui Nui (Figure 22), and Hawaii island (Figure 23). For the shallow habitat range (0-30 m), Uku hot spots represent 0% of the area, core habitats are 0.2%, basic EFH is 55.4%, and other (i.e., non-EFH habitats) are 44.4% of the area. For the deep habitat range (30-300 m), Uku hot spots are 0.09% of the area, core habitats are 2.4%, basic EFH is 59.8%, and other habitats are 37.3% of the area. Most of the hot spot and core habitats are in the deeper depth ranges of Penguin Bank (Figure 24) on the N-NW coasts of the islands, such as the Kohala coast of Hawaii Island (Figure 25). Basic EFH habitats cover the majority of both shallow and deep habitats. The “other” habitats (i.e., non-EFH) were predominately either shallow, nearshore habitats adjacent to shorelines or the habitats deeper than ~150 m throughout the main Hawaiian Islands.

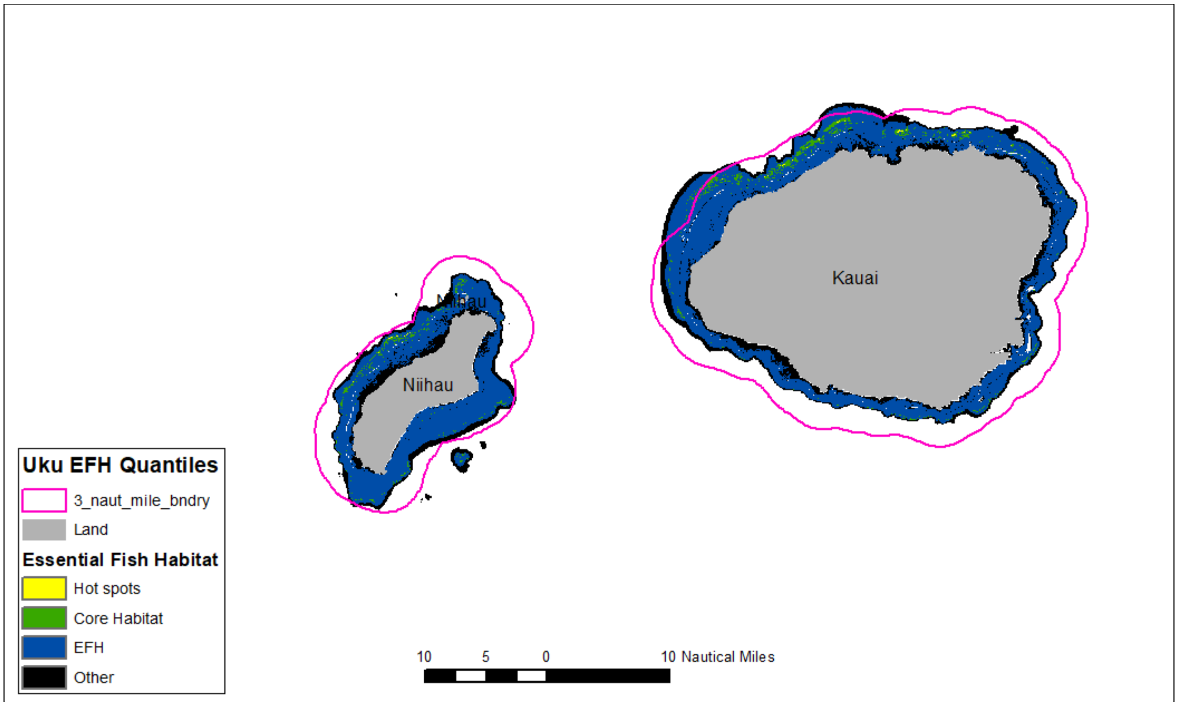


Figure 20. Uku (*Aprion virescens*) EFH hot spots (predicted occurrence probability of 0.75-1.0), core habitats (0.5-0.75), and basic EFH (0.05-0.5) around Kauai and Niihau from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

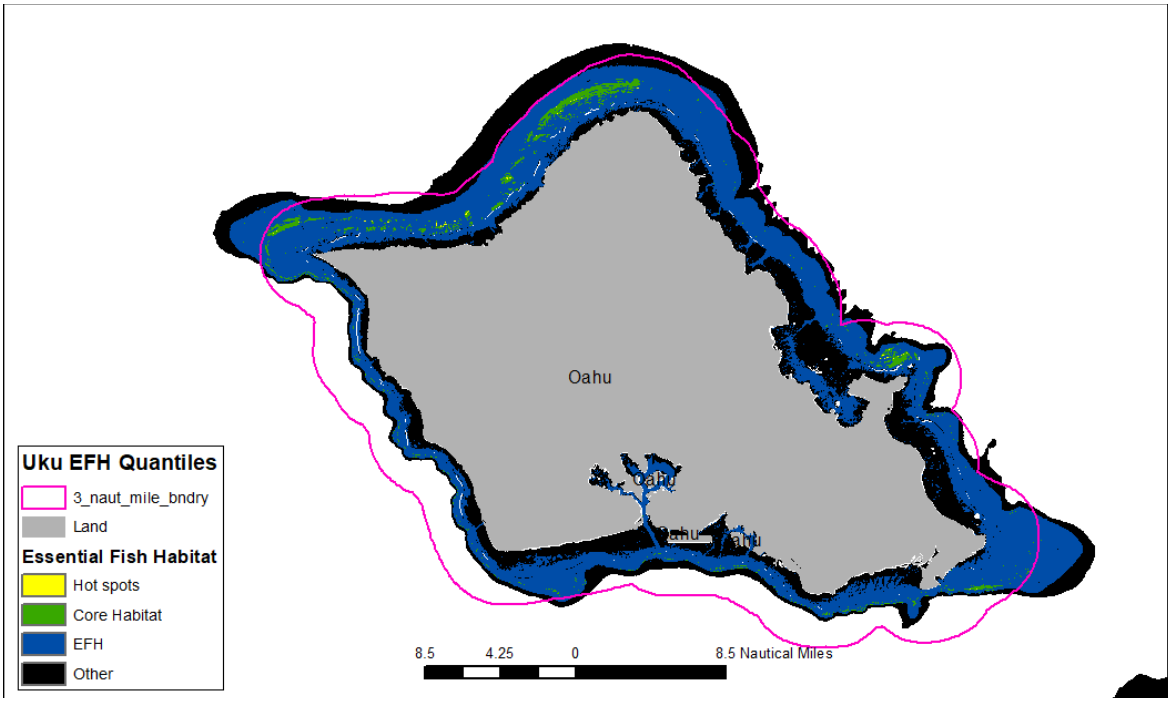


Figure 21. Uku (*Aprion virescens*) EFH hot spots (i.e., predicted occurrence probability of 0.75-1.0), core habitats (0.5-0.75), and basic EFH (0.05-0.5) around Oahu from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

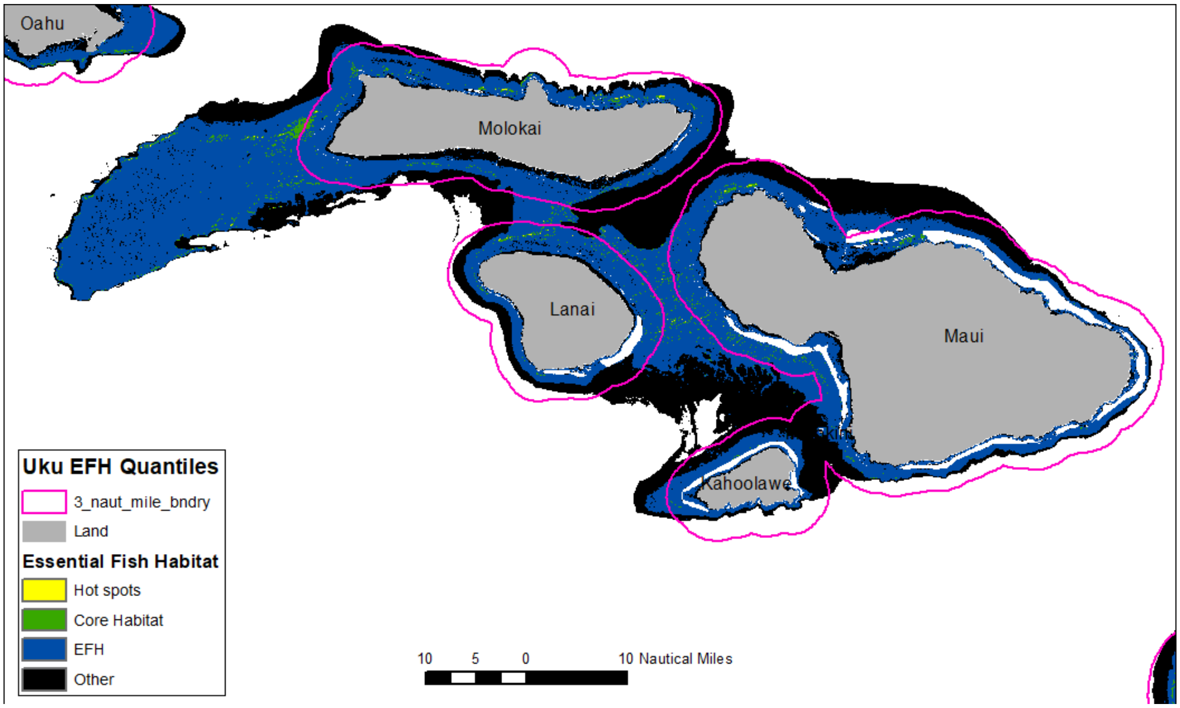


Figure 22. Uku (*Aprion virescens*) EFH hot spots (i.e., predicted occurrence probability of 0.75-1.0), core habitats (0.5-0.75), and basic EFH (0.05-0.5) around Maui Nui from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

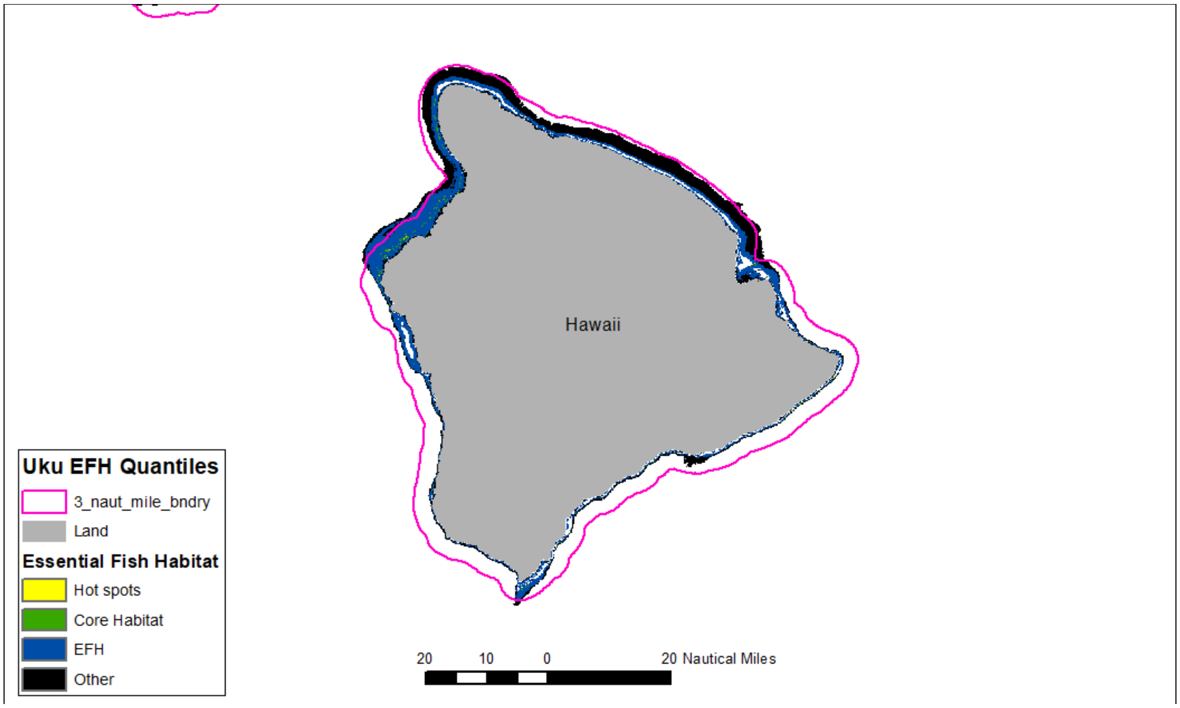


Figure 23. Uku (*Aprion virescens*) EFH hot spots (i.e., predicted occurrence probability of 0.75-1.0), core habitats (0.5-0.75), and basic EFH (0.05-0.5) around Hawaii island from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

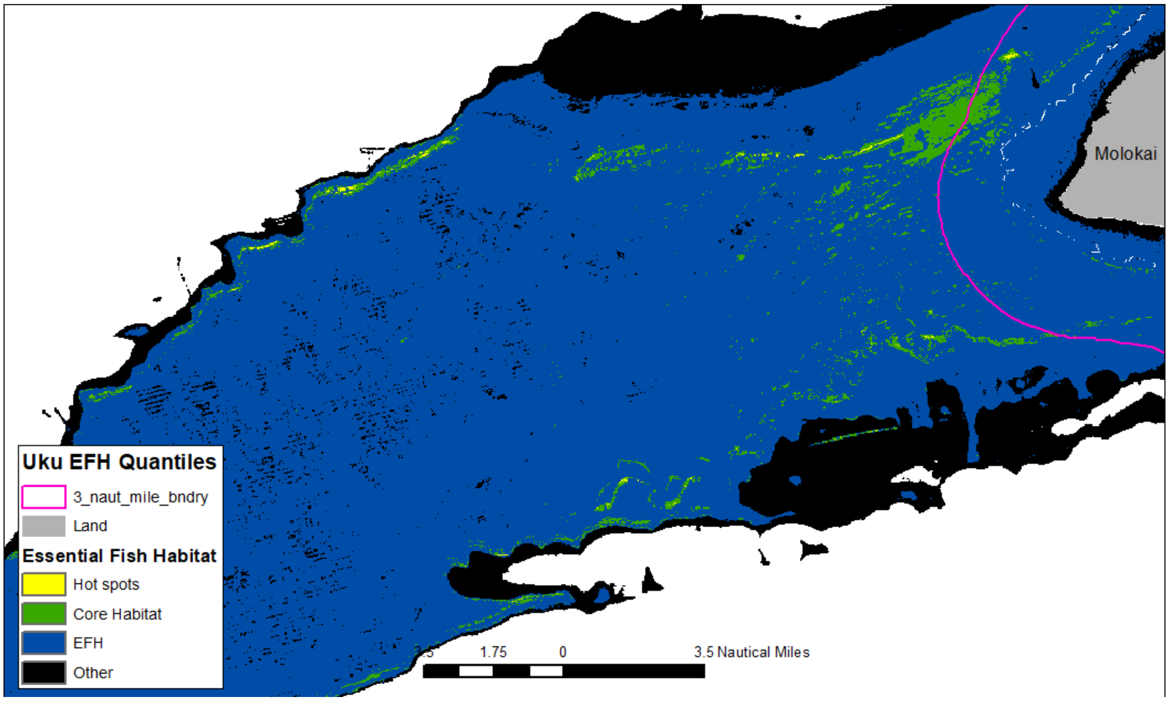


Figure 24. Uku (*Aprion virescens*) EFH hot spots (i.e., predicted occurrence probability of 0.75-1.0), core habitats (0.5-0.75), and basic EFH (0.05-0.5) on Penguin Bank from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

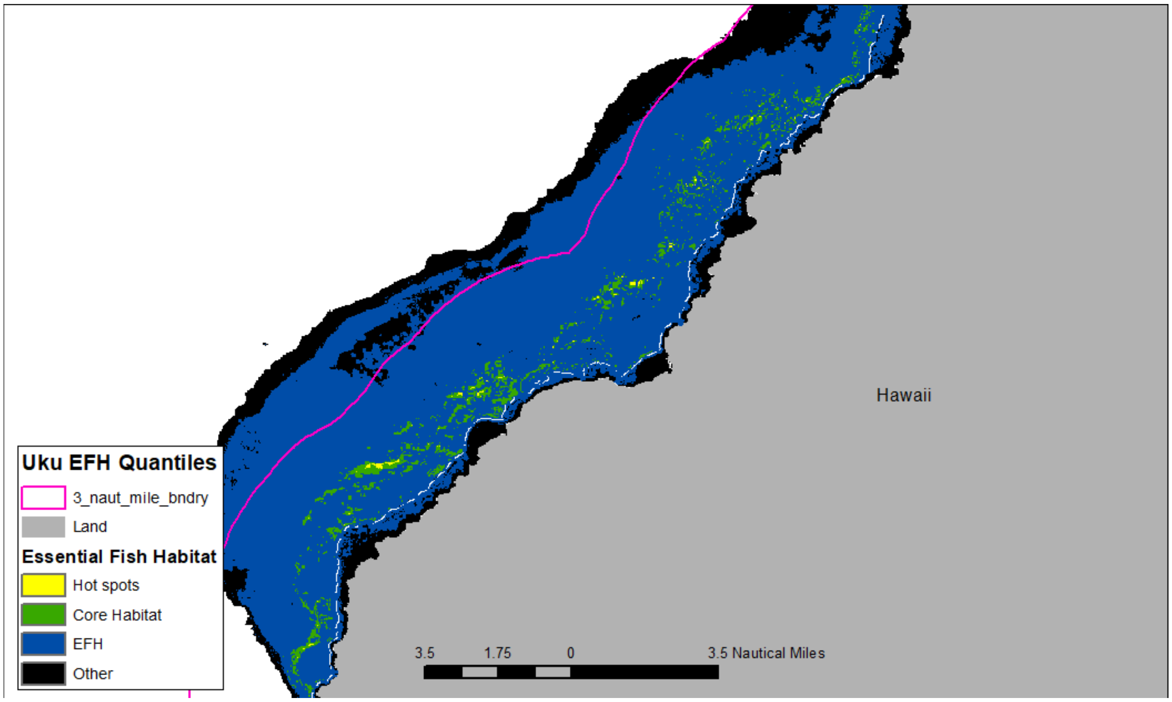


Figure 25. Uku (*Aprion virescens*) EFH hot spots (i.e., predicted occurrence probability of 0.75-1.0), core habitats (0.5-0.75), and basic EFH (0.05-0.5) on Penguin Bank from optimal BRT models for shallow (0-30 m) and deep (30-300 m) habitats.

CONCLUSIONS

A species distribution modelling approach was successfully demonstrated for the delineation of EFH for sub-adult/adult Uku (*Aprion virescens*) in the main Hawaiian Islands. The availability of fisheries-independent diver surveys for shallow (0 – 30 m) habitats and stereo-video camera surveys for deeper (30 – 300 m) provided sufficient observations to model the predicted occurrence of Uku across the entire spatial domain. Output from the SDMs were used to create maps delineating Uku EFH including continuous probability of occurrence maps as well as EFH categorical maps. The resulting categorical maps allowed the geographic identification of “hot spot”, “core habitat”, and “basic EFH” areas for Uku EFH to facilitate management activities.

Next steps to improve EFH delineations include an expansion of the environmental covariate set to include “dynamic” variables such as temperature, salinity, productivity, and current velocity. These may provide improved model fits and better allow forecasting of changes to EFH due to climate change effects. These will be attempted and documented in a follow-up report to this work. Another deficiency was the lack of sufficient fish length observations from the deep camera surveys necessary to model the sub-adult and adult fish independently. To address this issue, existing videos could be reanalyzed to collect Uku length data for this purpose. Any future video survey analysis from the BFISH project should also include Uku as a priority species to collect length data. These data, coupled with the existing information from the shallow diver surveys, should allow the EFH for the sub-adult (i.e., juvenile) and adult Uku populations to be modeled separately (although I anticipate the results to be similar as there is not an apparent ontogenetic shift in habitats for Uku in the MHI). Even with these potential improvements, the analyses in this report represent the first model-based approach to delineating EFH for a U.S. Western Pacific stock and could serve as a framework for the EFH descriptions of other managed species in the region.

References Cited

Amin R, Richards BL, Misa WFXE, Taylor JC, Miller DR, Rollo AK, Demarke C, Singh H, Young GC, Childress J, Ossolinski JE, Reardon RT, Koyanagi KH (2017) The Modular Optical Underwater Survey System. *Sensors* 17: 2309.

Asher J, Williams ID, Harvey ES (2017) An assessment of mobile predator populations along shallow and mesophotic depth gradients in the Hawaiian Archipelago. *Scientific Reports* 7: 3905 DOI:10.1038/s41598-017-03568-1

Battista TA, Costa BM, Anderson SM (2007) Shallow-water benthic habitats of the main eight Hawaiian Islands (DVD). NOAA Technical Memorandum NOS NCCOS 61, Biogeography Branch. Silver Spring, MD

Burrough, P.A., McDonnell, R.A., 1998. Principals of Geographic Information Systems (revised Edition). Clarendon Press, Oxford, pp. p.333.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802-813.

Fletcher CH, Bochicchio C, Conger CL, Engels MS and others (2008) Geology of Hawaii reefs. In: Riegl BM, Dodge RE (eds) Coral reefs of the USA. Springer, New York, NY

Franklin EC, Jokiel PL, Donahue MJ (2013) Predictive modeling of coral distribution and abundance in the Hawaiian Islands. *Mar. Ecol. Prog. Ser.* 481, 121–132.

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning: data mining, inference and prediction. Springer-Verlag, New York, NY

Heenan A, Williams ID, Acoba T, DesRochers A, Kosaki RK, Kanemura T, Nadon MO, Brainard RE (2017) Long-term monitoring of coral reef fish assemblages in the Western central Pacific. *Scientific Data* 4: 170176. DOI: 10.1038/sdata.2017.176

Hosmer, J.D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression. John Wiley & Sons, Inc., Hoboken, NJ p.528.

Merritt D, Donovan MK, Kelley C, Waterhouse L, Parke M, Wong K, Drazen JC (2011) BotCam: a baited camera system for nonextractive monitoring of bottomfish species. *Fishery Bulletin* 109: 56-67.

Oyafuso ZS, Drazen JC, Moore CH, Franklin EC (2017) Habitat-based species distribution modelling of the Hawaiian deepwater snapper-grouper complex. *Fisheries Research* 195: 19-27.

Pirtle J, Laman N, Thorson J (2020) Discussion paper on advancing essential fish habitat descriptions and maps for the 2022 5-year review. Report to North Pacific Regional Fishery Management Council. 150 pp.

Ridgeway G (2012) Generalized boosted regression models. Documentation on the R package ‘gbm’, version 2.18, www. r-project.org/

Schröder, B., Richter, O., 2000. Are habitat models transferable in space and time. J. Nat. Conserv. 8, 195–205.

Western Pacific Regional Fishery Management Council [WPRFMC] (2016) Amendment 4 to the Fishery Ecosystem Plan for the Hawaii Archipelago: Revised Descriptions of Essential Fish Habitat and Habitat Areas of Particular Concern for Bottomfish and Seamount Groundfish of the Hawaiian Archipelago. 243 pp.

Wright, D., Lundblad, E., Larkin, E., Rinehart, R., Murphy, J., Cary-Kothera, L., Draganov, K., 2005. ArcGIS Benthic Terrain Modeler (BTM), v. 3.0. Environmental Systems Research Institute, NOAA Coastal Services Center, Massachusetts Office of Coastal Zone Management.

Appendix: EFH Mapping Requirements

Paragraphs from the Code of Federal Regulations that are relevant to the geographic representations of EFH are excerpted below (source: NOAA NMFS Habitat Conservation EFH Mapper: https://www.habitat.noaa.gov/application/efhmapper/index.html) . The full text of this section of the CFR can be accessed here

(50 CFR Ch. VI (10-1-02 Edition)
Subpart J-Essential Fish Habitat

EFH)§ 600.815 Contents of Fishery Management Plans.

(a) Mandatory contents

(1) Description and identification of EFH

(i) Overview. FMPs must describe and identify EFH in text that clearly states the habitats or habitat types determined to be EFH for each life stage of the managed species. FMPs should explain the physical, biological, and chemical characteristics of EFH and, if known, how these characteristics influence the use of EFH by the species/life stage. FMPs must identify the specific geographic location or extent of habitats described as EFH. FMPs must include maps of the geographic locations of EFH or the geographic boundaries within which EFH for each species and life stage is found.

(iv) EFH determination.

(B) FMPs must describe EFH in text, including reference to the geographic location or extent of EFH using boundaries such as longitude and latitude, isotherms, isobaths, political boundaries, and major landmarks. If there are differences between the descriptions of EFH in text, maps, and tables, the textual description is ultimately determinative of the limits of EFH. Text and tables should explain pertinent physical, chemical, and biological characteristics of EFH for the managed species and explain any variability in habitat usage patterns, but the boundaries of EFH should be static.

(v) mapping requirements.

(A) FMPs must include maps that display, within the constraints of available information, the geographic locations of EFH or the geographic boundaries within which EFH for each species and life stage is found. Maps should identify the different types of habitat designated as EFH to the extent possible. Maps should explicitly distinguish EFH from non-EFH areas. Councils should confer with NMFS regarding mapping standards to ensure that maps from different Councils can be combined and shared efficiently and effectively. Ultimately, data used for mapping should be incorporated into a geographic information system (GIS) to facilitate analysis and presentation.

(B) Where the present distribution or stock size of a species or life stage is different from the historical distribution or stock size, then maps of historical habitat boundaries should be included in the FMP, if known.

(C) FMPs should include maps of any habitat areas of particular concern identified under paragraph (a)(8) of this section.

Appendix: Species distribution modeling (SDM) software code

Example code in the R statistical software language used to perform species distribution modeling for Uku EFH delineation in the main Hawaiian Islands.

```
#######################################
## Aprion virescens Species Distribution Modeling
## Erik Franklin, Zack Oyafuso
## Boosted Regression Trees
#######################################
setwd("") #need to set working directory

###########################
## Import Libraries
###########################
library(dismo); library(gbm)
library(PresenceAbsence)
library(corrplot)
source("brt.functions.R")

###########################
## Import Data
###########################
uku_data = read.csv("") # need to set data file

#############################
## correlation matrix for covariates
#############################
corrplot(cor(uku_data[,]),
      method = "number",
      type = "upper" # show only upper
      )

##############################
## For BRTs, test BRT heuristic using different three types of model settings
## All records are used for this section
## bag: bagging fraction (50% or 75%)
## tcomp: tree complexity (2, 3, 4, or 5)
## lrs: learning rate (0.001, 0.002, 0.003, 0.004, or 0.005)
##############################
spp_name <- "Aprvire"
brt_settings <-  expand.grid(species = spp_name,
                   bag = c(0.75, 0.50),
                   tcomp = 2:5,
                   lrs = c(0.001, 0.005, 0.01),
                   ntrees = NA, #Number of trees
```

```
                    devmean = NA, #mean cv deviance
                    stringsAsFactors = FALSE)


# split datasets into 70% train and 30% test (i.e., independent)
splitdf <- function(dataframe, seed=NULL) {
  if (!is.null(seed)) set.seed(seed)
  index <- 1:nrow(dataframe)
  trainindex <- sample(index, trunc(length(index)*0.7))
  trainset <- dataframe[trainindex, ]
  testset <- dataframe[-trainindex, ]
  list(trainset=trainset,testset=testset)
}

uku_splits <- splitdf(uku_data, seed=105)
uku_train <- uku_splits$trainset
uku_test <- uku_splits$testset
dim(uku_train); dim(uku_test); dim(uku_data)

for (i in 1:nrow(brt_settings)) {
	temp_fit = dismo::gbm.step(data = uku_train,
				#Column ids of covariates in argument data
				gbm.x = c(
					#4, #lat
					#5, #long
					6, #depth
					7, #aspect
					8, #sand
					9, #rugosity
					10), #slope
				#name of column that contains response
				gbm.y = spp_name,
				family = "bernoulli",
				tree.complexity = brt_settings$tcomp[i],
				learning.rate = brt_settings$lrs[i],
				bag.fraction = brt_settings$bag[i],
				verbose = TRUE)

	brt_settings[i, c("ntrees", "devmean")] <-
		c(temp_fit$n.trees, temp_fit$cv.statistics$deviance.mean)
}

############################
## Trim settings that used less than 1500 trees
## Choose settings with the lowest mean deviance
###########################
```

```
brt_settings_trimmed <- subset(x = brt_settings,
                               subset = ntrees > 1500)

(best_settings <- brt_settings_trimmed[which.min(brt_settings_trimmed$devmean), ])

##############################
## Setup up training models with the optimized BRT settings
#############################
train_uku = gbm.step(data = uku_train,
                     gbm.x = c(6:10),
                     gbm.y = spp_name,
                     family = "bernoulli",
                     tree.complexity = best_settings$tcomp,
                     learning.rate = best_settings$lrs,
                     bag.fraction = best_settings$bag)

##############################################
## Parsimonous Models on Test Dataset
##############################################
test_uku = gbm.step(data = uku_test,
                    gbm.x = c(6:10),
                    gbm.y = spp_name,
                    family = "bernoulli",
                    tree.complexity = best_settings$tcomp,
                    learning.rate = best_settings$lrs,
                    bag.fraction = best_settings$bag)

###################################
## Evaluation Metrics
###################################
AUC_return = function(model.name, sp_code, thres) {

        preds <- predict.gbm(model.name,
                             uku_test,
                             n.trees=model.name$gbm.call$best.trees,
                             type="response")

        d <- cbind(uku_test[,sp_code], preds)
        pres <- d[d[,1]==1, 2]
        abs <- d[d[,1]==0, 2]
        e <- evaluate(p=pres, a=abs, tr = thres)
        return(round(e@auc,2))

}

output_df = data.frame(species = spp_name)
```

```
#output_df$Raw_I <- round(ape::Moran.I(x = uku_test[, spp_name], w =
test.dists.inv)$observed, 3)
test_pred <- predict.gbm(object = train_uku,
                     newdata = uku_test,
                     n.trees = train_uku$gbm.call$best.trees,
                     type = "response")
#residual = test_pred - uku_test[, spp_name]

#output_df$Environ_I =  round(ape::Moran.I(x = residual, w = test.dists.inv)$observed, 3)

tv = sum(uku_data[, spp_name]) / nrow(uku_data)
name_assign = paste('cmx_', spp_name, sep = '')
assign(name_assign,
      PresenceAbsence::cmx(cbind(1:nrow(uku_test), uku_test$Aprvire,
                            predict.gbm(train_uku,
                                      uku_test,
                                      type = 'response',
                                      n.trees = train_uku$n.trees)), threshold = tv))

output_df$Environ_DEV = 1 - (test_uku$self.statistics$mean.res /
                              test_uku$self.statistics$mean.null)
output_df$AUC = AUC_return(train_uku, spp_name, tv)
output_df$TSS = PresenceAbsence::sensitivity(get(name_assign), st.dev = FALSE) +
PresenceAbsence::specificity(get(name_assign), st.dev = FALSE) - 1
output_df$kappa = PresenceAbsence::Kappa(get(name_assign), st.dev = FALSE)
output_df$specific = PresenceAbsence::specificity(get(name_assign), st.dev = FALSE)
output_df$sens = PresenceAbsence::sensitivity(get(name_assign), st.dev = FALSE)
output_df

########################
## Percent Variable Importance
########################
train_uku$contributions


########################
## Plot fitted functions
########################
gbm.plot(train_uku)#, write.title=TRUE)

gbm.plot.fits(train_uku)

#########################
## Interactions
#########################
find.int = gbm.interactions(train_uku)
```

```
find.int$interactions
find.int$rank.list

## plot largest interaction size
gbm.perspec(train_uku, 3, 1)

##########################
## Predict to grids
##########################

eval.data <- read.csv("deep_covars.csv", as.is=T)
names(eval.data) = c("FID", "pointid", "depth", "sand", "surf_ratio", "slope", "aspect",
"rugosity","lat","long")
eval.data = eval.data[ , c("lat", "long", "depth", "aspect", "sand", "rugosity", "slope")]

gbm.predict.grids(train_uku, eval.data, want.grids = FALSE, sp.name = "aprvire_pred")

aprvire.out = cbind(eval.data, aprvire_pred)
write.csv(aprvire.out,"pred\\aprvire_deep_total.csv")
```