Marine and Fishery Sciences 39 (4): xxx-xxx (2026)
https://doi.org/10.47193/mafis.39420260011003

ORIGINAL RESEARCH

Random Forest habitat suitability modeling for *Acanthurus* spp. in Philippine coral reefs: environmental drivers and climate projections

Nelson C. Perante*

College of Fisheries and Marine Sciences, Samar State University - Mercedes Campus, Mercedes, Samar, Eastern Visayas, Philippines.
ORCID *Nelson C. Perante* https://orcid.org/0000-0002-2156-8354

Marine and Fishery Sciences MAFIS

ABSTRACT. Surgeonfish of the genus *Acanthurus* (family Acanthuridae) are ecologically pivotal herbivores on Indo-Pacific coral reefs, yet their habitat requirements and distributional responses to environmental gradients and climate change remain poorly characterized in the Philippines. A Random Forest (RF) machine learning framework was used to model habitat suitability for six commercially and ecologically important species –*A. lineatus*, *A. nigricans*, *A. olivaceus*, *A. triostegus*, *A. xanthopterus*, and *A. pyroferus*– across Philippine coral reef systems. Geo-referenced occurrence records (n = 352-714 per species) from OBIS, GBIF, and ReefBase were paired with ten environmental predictors encompassing reef structure, oceanographic conditions, bathymetry, and physical disturbance. RF models achieved mean AUC of 0.929 and TSS of 0.805, indicating excellent predictive performance. Coral cover, depth, and rugosity were the dominant habitat predictors across species. Current suitable habitat ranged from 6,890 km^2 (*A. olivaceus*) to 11,350 km^2 (*A. xanthopterus*). Under SSP5-8.5 by 2100, projected habitat declined by 25.4-36.1%, driven primarily by thermally induced coral cover loss. *Acanthurus triostegus* was most vulnerable (-36.1%) and *A. xanthopterus* most resilient (-25.4%). Multi-species overlap analysis identified Verde Island Passage, Tubbataha Reef, and Apo Island as priority conservation nodes. Gap analysis revealed three unprotected high-overlap zones as candidates for new marine protected area (MPA) establishment. These findings provide a spatial evidence base for climate-adaptive reef fisheries management across the Indo-Pacific.

Key words: Random Forest, habitat suitability, Acanthuridae, CMIP6, marine protected areas.

*Correspondence:
nelson.perante@ssu.edu.ph

Received: 5 June 2026
Accepted: 13 July 2026

ISSN 2683-7595 (print)
ISSN 2683-7951 (online)

https://ojs.inidep.edu.ar

Journal of the Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP)

Modelado de idoneidad de hábitat mediante Bosque Aleatorio para *Acanthurus* spp. en arrecifes de coral filipinos: factores ambientales y proyecciones climáticas

RESUMEN. Los peces cirujano del Género *Acanthurus* (Familia Acanthuridae) son herbívoros clave desde el punto de vista ecológico en los arrecifes de coral del Indo-Pacífico. Sin embargo, sus requerimientos de hábitat y respuestas distributivas a los gradientes ambientales y al cambio climático aún no se conocen bien en Filipinas. Se utilizó un modelo de aprendizaje automático basado en Bosque Aleatorio (RF) para modelar la idoneidad de hábitat de seis especies de importancia comercial y ecológica: *A. lineatus*, *A. nigricans*, *A. olivaceus*, *A. triostegus*, *A. xanthopterus* y *A. pyroferus*, en los sistemas de arrecifes de coral filipinos. Los registros de presencia georreferenciados (n = 352-714 por especie) de OBIS, GBIF y ReefBase se combinaron con diez predictores ambientales que abarcan la estructura del arrecife, las condiciones oceanográficas, la batimetría y las perturbaciones físicas. Los modelos RF lograron un AUC medio de 0,929 y un TSS de 0,805, lo que indica un excelente rendimiento predictivo. La cobertura de coral, la profundidad y la rugosidad fueron los predictores de hábitat dominantes en todas las especies. El hábitat adecuado actual varió de 6.890 km^2 (*A. olivaceus*) a 11.350 km^2 (*A. xanthopterus*). Bajo SSP5-8.5 para 2100, el hábitat proyectado disminuyó en 25,4-36,1%, impulsado principalmente por la pérdida de cobertura de coral

inducida térmicamente. *Acanthurus triostegus* fue la más vulnerable (-36,1%) y *A. xanthopterus* la más resiliente (-25,4%). El análisis de superposición de múltiples especies identificó el Paso de la Isla Verde, el Arrecife Tubbataha y la Isla Apo como nodos de conservación prioritarios. El análisis de brechas reveló tres zonas de alta superposición no protegidas como candidatas para el establecimiento de nuevas Áreas Marinas Protegidas (AMP). Estos hallazgos proporcionan una base de evidencia espacial para la gestión de pesquerías de arrecife adaptadas al clima en todo el Indo-Pacífico.

Palabras clave: Bosque Aleatorio, idoneidad del hábitat, Acanthuridae, CMIP6, áreas marinas protegidas.

INTRODUCTION

Coral reefs are among the most biodiverse and economically productive ecosystems on Earth, supporting approximately 25% of all marine species while occupying less than 0.2% of the ocean surface (Reaka-Kudla 1997; Cesar et al. 2003). In the Philippines, located at the apex of the Coral Triangle, the global center of marine biodiversity, reefs cover an estimated 26,000 km^2 and sustain the livelihoods of over 1.5 million small-scale fishers, contributing approximately USD 1.4 billion annually to the national economy through fisheries, tourism, and coastal protection services (Burke et al. 2011; BFAR 2023). The Philippine Archipelago hosts over 2,400 reef fish species, including at least 11 species of the genus *Acanthurus*.

Surgeonfish (family Acanthuridae) occupy a functional role on coral reefs second only to corals themselves in determining ecosystem state. As herbivores, grazers, and detritivores, *Acanthurus* species regulate benthic community composition by preventing macroalgal dominance, facilitating coral recruitment, and redistributing benthic organic matter through bioerosion and bioturbation (Bellwood et al. 2004; Hughes et al. 2007). The removal of herbivorous fish assemblages through overfishing is well documented as a key precursor to phase shifts from coral to algal dominance on disturbed reefs (Hughes 1994; Mumby et al. 2007). In the Philippines specifically, catch-trend analyses across small-scale municipal reef fisheries indicate that most surveyed fishing grounds are overexploited or approaching collapse, with herbivore-targeting gear among the dominant contributing effort (Muallil et al. 2014). Field surveys further show that reef-wide macroalgal cover remains low only when total herbivorous fish biomass exceeds approximately 1.0 t km^{-2} and parrotfish biomass exceeds 2.0-4.0 t km^{-2}. This indicates a narrow biomass margin within which Philippine reef herbivore assemblages, including *Acanthurus*, currently operate (Muallil et al. 2020). In the Philippines, where reef fishing pressure is exceptionally high, maintaining viable surgeonfish populations is a first-order conservation priority.

Despite this ecological importance, spatial information on the habitat requirements and distributional limits of Philippine *Acanthurus* populations is sparse. Existing distributional records are dominated by museum specimens and opportunistic sightings, supplemented by localized survey data from high-profile reef areas. This knowledge gap is particularly acute for species such as *A. olivaceus* (Rüppell, 1829) and *A. pyroferus* (Kittlitz, 1834), whose habitat associations outside of well-studied central Philippine reef systems remain poorly described.

Machine learning-based species distribution models (SDMs) offer a powerful approach for addressing this gap, particularly Random Forest (RF), which has emerged as one of the most robust methods for ecological prediction (Prasad et al. 2006; Cutler et al. 2007). RF is a non-parametric ensemble method that constructs large numbers of decision trees on bootstrapped training subsets and aggregates their predictions, without requiring a priori assumptions about species-environment relationships (Breiman 2001). For coral reef systems, RF has been applied successfully to coral cover prediction

(González-Rivero et al. 2020), fish biomass mapping (Mellin et al. 2012), and MPA effectiveness assessment. Deep learning architectures (e.g. convolutional and graph neural networks) are increasingly explored in marine species distribution modelling and can capture complex spatio-temporal covariate interactions, but they typically require substantially larger and denser occurrence datasets than are available for most Philippine reef fish, and they offer less directly interpretable variable-importance output than tree-based ensembles (Klaassen et al. 2025). Given the comparatively modest per-species sample sizes available for Philippine *Acanthurus* (n = 352-714) and the emphasis on interpretable, ecologically meaningful predictor rankings, RF was selected as the more appropriate ensemble approach for this study, consistent with its wide use in other reef fish and coral cover modelling applications (Pickens et al. 2021). Its capacity to handle non-linear interactions, resist overfitting, and provide robust variable importance estimates makes it particularly well-suited to reef fish habitat modeling. A systematic review of 225 marine fish distribution studies found that guild- and genus-level habitat models consistently identify transferable structural and oceanographic predictors for coral-reef-associated fishes, supporting the multi-species, genus-focused RF framework adopted here rather than a single-species approach (Pickens et al. 2021).

Climate change is rapidly altering the environmental template of Philippine reefs through ocean warming, acidification, and shifts in typhoon and ENSO seasonality. Mass coral bleaching events have become increasingly frequent, with the 2015-2016 and 2022-2023 bleaching events causing documented declines in *Acanthurus* abundance in thermally stressed reef zones (Pratchett et al. 2011; Emslie et al. 2018). Projections under high-emissions scenarios suggest thermally suitable reef habitat will decline by 30-50% in Philippine waters by 2100 (Logan et al. 2014), with cascading implications for reef fish communities.

This study develops the first comprehensive, multi-species RF habitat suitability framework for Philippine *Acanthurus*, linking current distributional patterns to reef structural, oceanographic, and bathymetric predictors and projecting future habitat availability under three SSP climate scenarios. The specific objectives were to: (1) compile and quality-control geo-referenced occurrence records for six target species; (2) develop and evaluate species-specific RF habitat suitability models using ten environmental predictors; (3) identify key environmental drivers of habitat suitability; (4) map current suitable habitat, identify multi-species priority zones, and conduct a gap analysis against existing MPA coverage to flag unprotected priority zones; and (5) project future habitat suitability under SSP1-2.6, SSP2-4.5, and SSP5-8.5 for 2050 and 2100 (representing low, intermediate, and high radiative-forcing pathways of approximately + 2.6, + 4.5, and + 8.5 W m^{-2} by 2100 under the CMIP6 Shared Socioeconomic Pathway framework, evaluated at the 2041-2060 [‘2050’] and 2081-2100 [‘2100’] 20-year mean horizons) to assess species vulnerability and inform climate-adaptive management. The methodological framework and distributional outputs developed here are directly applicable to SDM studies of reef-associated fish across the broader Indo-Pacific and Coral Triangle region.

MATERIALS AND METHODS

Study area

The study area encompasses the Philippine Archipelago and surrounding marine waters (approximately 5° N-21° N, 116° E-128° E), covering an estimated 965,000 km^2 of the western Pacific Ocean including the South China Sea, Sulu Sea, Celebes Sea, and Philippine Sea margins (Figure 1). The Philippines contains over 7,600 islands with a combined coastline of approximately 36,290 km (BFAR 2023). Spatial boundaries for all analyses followed the Philippine archipelagic baselines defined under Republic Act No. 9522 (2009), which

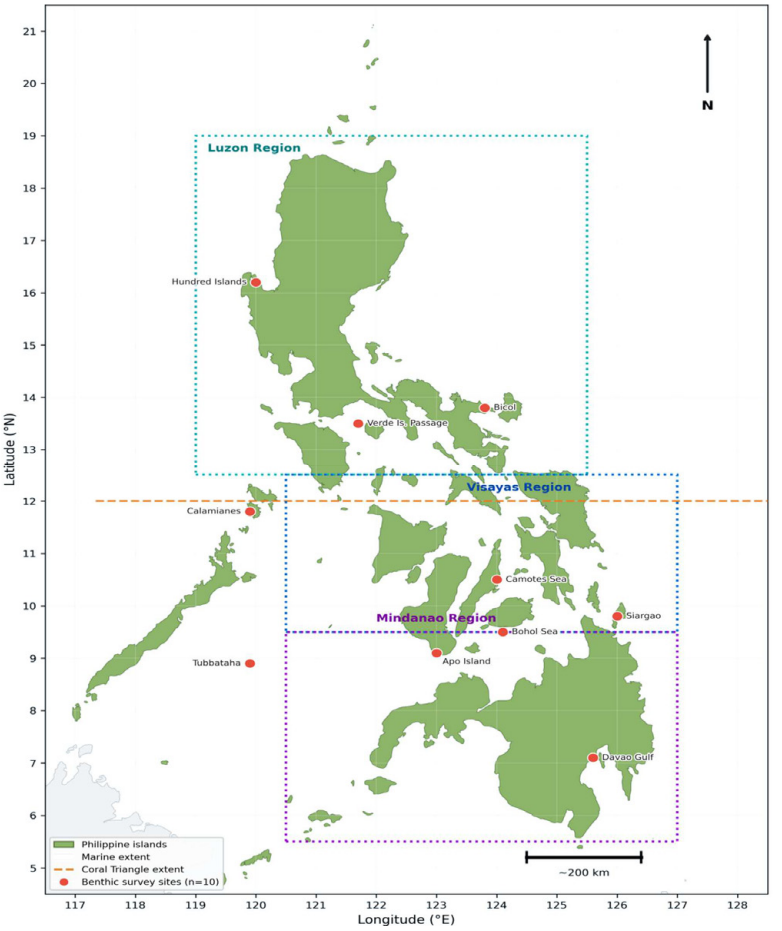

Figure 1. Study area showing the Philippine Archipelago, benthic survey sites used for *Acanthurus* occurrence data compilation (red circles, n = 10 primary survey areas), three biogeographic analysis regions (dotted colored boxes), and the Coral Triangle boundary (dashed orange line). Inset: Pacific Ocean context.

delimits the country's territorial sea and archipelagic waters in accordance with the UN Convention on the Law of the Sea (https://www.namria.gov.ph/Downloads), hosting reef systems across three principal biogeographic zones: the Luzon Region (northern Philippines, including Verde Island Passage, the global center of marine shore fish diversity; Carpenter and Springer 2005), the Visayas Region (central Philippines), and the Mindanao Region (southern Philippines). The Tubbataha Reef National Park (Sulu Sea, 97,030 ha), a UNESCO World Heritage Site, is the country's largest no-take MPA.

Philippine reefs experience tropical oceanic conditions with mean annual sea-surface temperature (SST) of 27-30 °C and pronounced seasonal SST cycles (± 2-3 °C). Intra-annual photosynthetically active radiation (PAR) variability is substantial (12-47 mol photons m^{-2} d^{-1}), and reef exposure ranges from highly exposed oceanic fore-reefs to sheltered lagoonal backreefs. These baseline values were derived from the same 2010-2022 CoralTemp/NOAA ERDDAP (https://coralreefwatch.noaa.gov/product/5km/index_5km_sst.php) and MODIS-Aqua (https://nsidc.org/data/modis) climatologists used to build the SST and PAR predictor layers.

Species selection and occurrence data

Six surgeonfish species (*A. lineatus, A. nigricans, A. olivaceus*, *A. triostegus*, *A. xanthopterus*, and *A. pyroferus*) were selected based on their ecological importance as functional herbivores and detritivores, commercial significance in artisanal reef fisheries, and sufficient occurrence record availability, consistent with established criteria for selecting focal taxa in reef fish habitat modelling based on functional role, fisheries relevance, and data sufficiency (Pickens et al. 2021). Occurrence records were downloaded from the Ocean Biodiversity Information System (OBIS; obis.org), the Global Biodiversity Information Facility (GBIF; gbif.org), and ReefBase (reefbase.org) in January 2024. Records were filtered to retain only those within Philippine waters (bounding box: 4° N-22° N, 114° E-130° E) with geo-referenced coordinates, collection dates after 1975, and coordinate uncertainty < 10 km. The 1975 cutoff excludes pre-satellite-era records for which reef habitat conditions can no longer be cross-checked against remotely sensed baseline data, while the 10 km coordinate-uncertainty threshold was set to be smaller than the coarsest environmental predictor grain used here (~ 4-5 km chlorophyll-a and PAR layers), so that occurrence records could be reliably matched to a single predictor cell.

Duplicate records at identical coordinates and records with implausible depth values (> 200 m) were removed. Spatial thinning was applied using a minimum nearest-neighbor distance of 5 km via the spThin package in R (Aiello-Lammens et al. 2015) to reduce spatial autocorrelation and sampling bias effects on model performance. For each species, 10,000 background pseudo-absence points were drawn using a target-group sampling approach constrained to reef-containing cells (within 2 km of ReefBase reef polygons; Phillips et al. 2009). This 2 km ReefBase reef-polygon buffer served as the coral reef extent mask applied throughout the study: the same buffer both constrained pseudo-absence generation here and restricted the final habitat suitability prediction surfaces (see Habitat suitability mapping below), ensuring that background points, training data, and prediction outputs were confined to a consistent reef-associated spatial domain. Final post-thinning occurrence counts ranged from n = 352 (*A. pyroferus*) to n = 714 (*A. triostegus*).

Environmental predictors

Ten environmental predictor variables were selected based on their established ecological relevance to reef fish habitat selection, foraging ecology, and physiological tolerances These ten predictors span the four covariate classes most consistently linked to coral reef fish distribution in the broader SDM literature –benthic structure, thermal/oceanographic conditions, bathymetry, and physical disturbance– following the predictor taxonomy recommended for marine fish habitat models (Pickens et al. 2021) (Tables 1 and 2; Figure 2). Multi-year climatological means (2010-2022) –chosen because it is the longest continuous, cloud-gap-filled period available from MODIS-Aqua Ocean Color and CoralTemp/NOAA ERDDAP at the time of model development, and because it post-dates the 2010 mass bleaching event, avoiding conflation of pre- and post-disturbance reef states within a single averaging window– were computed for all time-varying predictors. Raster layers were resampled to a common 500 m resolution using bilinear interpolation and clipped to the study extent. Bilinear interpolation was used rather than nearest-neighbour resampling to avoid abrupt step artefacts at coarse-to-fine transitions, following common practice in reef SDM raster preparation; however, we note that this procedure smooths rather than adds true fine-scale information, and predictor layers originating from coarser native resolutions (e.g. ~ 4-5 km chlorophyll-a, PAR, and SST products) should be interpreted at their native ecological grain rather than at the nominal 500 m analysis resolution, a common source of over-interpretation in resampled species distribution models

Table 1. Target *Acanthurus* species, occurrence record counts, and ecological attributes. LC = Least Concern (IUCN Red List 2023).

Species	Common name	Records (n)	Depth range (m)	Diet guild	IUCN status	Primary reef zone
A. lineatus	Lined surgeonfish	623	1-15	Territorial herbivore	LC	Crest/flat
A. nigricans	Whitecheek surgeonfish	498	1-25	Browser	LC	Slope
A. olivaceus	Orangespot surgeonfish	441	3-46	Detritivore	LC	Slope/flat
A. triostegus	Convict surgeonfish	714	0.5-90	Grazer	LC	Flat/lagoon
A. xanthopterus	Yellowfin surgeonfish	389	1-90	Omnivore	LC	Slope/lagoon
A. pyroferus	Mimic surgeonfish	352	4-60	Browser	LC	Slope

Records are post-thinning dataset sizes. Depth ranges compiled from FishBase and ReefBase.

Table 2. Environmental predictor variables used in Random Forest habitat suitability models. PAR = Photosynthetically Active Radiation; GCRMN = Global Coral Reef Monitoring Network; GEBCO = General Bathymetric Chart of the Oceans; CMEMS = Copernicus Marine Environment Monitoring Service.

Predictor variable	Source	Resolution	Value range	Ecological rationale
Coral cover (%)	ReefBase/GCRMN/ local surveys	500 m	2.1-84.7	Primary structural habitat determinant
Depth (m)	GEBCO 2022	15 arc-sec	0-120	Drives light, pressure, community zonation
SST mean (°C)	CoralTemp/NOAA ERDDAP	~ 5 km	25.3-30.9	Thermal physiology and metabolic rates
SST range (°C)	CoralTemp climatology	~ 5 km	1.2-5.8	Thermal stress exposure index
Chlorophyll-a (mg m^{-3})	MODIS-Aqua Level 3	~ 4 km	0.04-1.83	Productivity/turbidity proxy
Rugosity index	Multi-beam bathymetry/GEBCO	500 m	1.01-3.94	Structural complexity; refuge availability
Distance to reef edge (km)	ReefBase/ UNEP-WCMC	100 m	0-18.4	Larval supply and connectivity
PAR (mol m^{-2} d^{-1})	MODIS-Aqua	~ 4 km	12.3-46.8	Photosynthetic productivity; algal food base
Wave energy index (kJ m^{-2})	ERA5/ECMWF; GEBCO-adjusted	0.25°	0.3-94.7	Physical disturbance threshold
Salinity	CMEMS GLORYS12 reanalysis	1/12°	30.1-34.8	Osmoregulatory constraint

(Sillero and Barbosa 2021). Collinearity was assessed using the Variance Inflation Factor (VIF); predictors with pairwise Pearson's $|r| > 0.80$ were evaluated for exclusion. The highest correlation observed was between coral cover and rugosity ($r = 0.61$, $p < 0.001$) and between depth and PAR

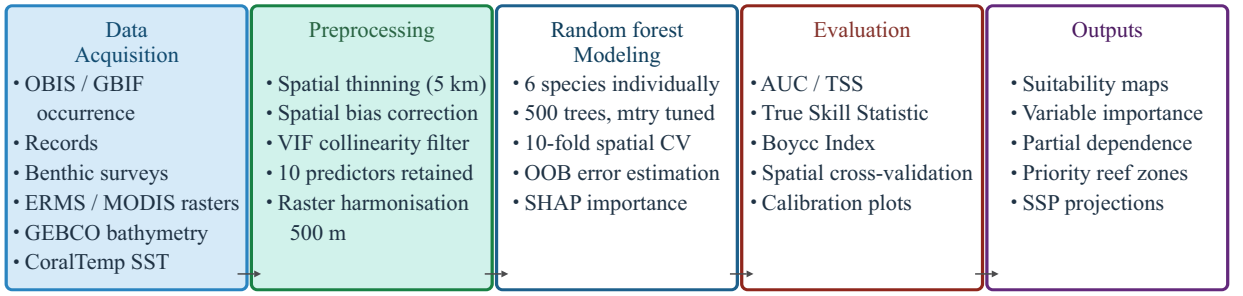

Figure 2. Methodological workflow for Random Forest habitat suitability modeling of *Acanthurus* spp. in Philippine coral reef systems, from data acquisition through spatial outputs.

($r = -0.54$, $p < 0.001$), both below the exclusion threshold.

Random Forest modeling

Model setup and hyperparameter tuning

Random Forest models were implemented in R using the randomForest package (Liaw and Wiener 2002) and the biomod2 ensemble framework (Thuiller et al. 2009). For each species, the full occurrence dataset was split 75:25 (training:testing) using a spatially stratified blocking approach implemented via the blockCV package (Valavi et al. 2019) with 100 km × 100 km spatial blocks, to avoid spatial autocorrelation between training and evaluation data. The number of trees (ntree) was set to 500 following standard convergence criteria, i.e. the point beyond which out-of-bag error stabilises and additional trees primarily reduce prediction variance rather than bias (Breiman 2001); 500 trees is also consistent with recommendations that the number of trees be set generously high rather than tuned, since more trees are never harmful to predictive accuracy and error curves for classification RF typically plateau well below this value (Probst and Boulesteix 2017). The number of variables randomly sampled at each split (mtry) was optimized via grid search over values 1-4 using 10-fold spatial cross-validation, with mean cross-validated AUC as the selection criterion. Minimum node size was set to 5 for most species and 7 for *A. xanthopterus* (lower sample size relative to training extent). Optimal hyperparameters and Out-of-Bag (OOB) error rates are summarized in Table 3.

Variable importance and partial dependence

Variable importance was quantified using the Mean Decrease in Gini Impurity (MDG) index. Permutation importance was calculated as a validation check. Partial dependence plots (PDPs) were generated for the four most important predictors of each species using the pdp package in R (Greenwell 2017), showing the marginal effect of each predictor on predicted habitat suitability while averaging over the distributions of all other predictors.

SHAP values

SHapley Additive exPlanations (SHAP) values were computed using the kernelshap R package (Mayer et al. 2023) to provide individual prediction-level explanations. SHAP values decompose each model prediction into contributions from individual predictors, complementing the global MDG importance summaries.

Model evaluation

Model predictive performance was evaluated on the 25% spatially blocked holdout partition using: (1) the Area Under the ROC Curve (AUC; Hanley and McNeil 1982), where values > 0.90 indicate excellent discrimination; (2) the True Skill Statistic (TSS; Allouche et al. 2006), with values > 0.70 considered good; (3) the Boyce Continuous

Table 3. Random Forest tuning parameters and performance metrics for six *Acanthurus* species. OOB = Out-of-Bag. mtry = number of variables randomly sampled at each split.

Species	Optimal ntree	Optimal mtry	Node size	OOB error (%)	Training AUC	Test AUC
A. lineatus	500	3	5	8.3	0.968	0.941
A. nigricans	500	3	5	9.1	0.954	0.927
A. olivaceus	500	2	5	9.8	0.948	0.919
A. triostegus	500	3	5	8.7	0.961	0.934
A. xanthopterus	500	2	7	10.4	0.941	0.912
A. pyroferus	500	3	5	8.4	0.966	0.938

All AUC values based on spatially blocked holdout partitions (blockCV 100 km).

Index (Hirzel et al. 2006); (4) Cohen's Kappa; and (5) model calibration assessed by comparing predicted probabilities against observed prevalence in decile-binned groups.

Metrics were calculated as follows. AUC integrates the ROC curve of true positive rate (sensitivity) against false positive rate (1 - specificity) across all classification thresholds. The TSS = sensitivity + specificity - 1, ranging from -1 to 1, with 0 indicating performance no better than random. Cohen's Kappa = $(P_o - P_e)/(1 - P_e)$, where P_o is observed proportional agreement between predicted and observed classes and P_e is the agreement expected by chance given the marginal class frequencies. The Boyce Continuous Index is the Spearman rank correlation between the predicted-to-expected (P/E) ratio of occurrences across suitability bins and the median suitability of each bin, ranging from -1 (counter-predictive) to 1 (perfect agreement with observed occurrence density).

Habitat suitability mapping

Habitat suitability was predicted across all reef-containing 500 m cells in the Philippine study area by applying trained RF models to the full raster predictor stack. For area calculations, cells were classified as suitable if the predicted suitability exceeded a threshold determined by maximizing TSS on the holdout partition (Liu et al. 2005). Suitability maps were restricted to cells within 2 km of known reef occurrence polygons from ReefBase.

Climate change projections

Future habitat suitability was projected under three SSP scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5) at two-time horizons (2041-2060 and 2081-2100) using a six-model CMIP6 ensemble (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, CNRM-CM6-1, CanESM5, ACCESS-ESM1-5). The SST, PAR, chlorophyll-a, and salinity were projected forward using delta change bias correction against the 2010-2022 baseline. Future coral cover was derived from a thermal bleaching threshold model (coral cover reduced proportionally as annual degree weeks exceed 8 °C-weeks; Logan et al. 2014). Rugosity, depth, distance to reef edge, and wave energy were held constant across projections. Projections were run separately for each species and CMIP6 model, and median ± inter-model range reported.

Multi-species overlap and MPA gap analysis

Multi-species overlap maps were generated by summing binary suitability layers across all six species. Cells classified as high-suitability for four or more species were designated high-priority zones. These were intersected with the World Da-

tabase on Protected Areas (WDPA) MPA layer for the Philippines to identify unprotected high-overlap zones representing conservation gaps. Spatial analysis was conducted in QGIS 3.32 and R.

RESULTS

Occurrence data and predictor collinearity

The final thinned occurrence datasets comprised 3,017 unique geo-referenced records across six species. *Acanthurus triostegus* had the largest dataset (n = 714), while *A. pyroferus* had the fewest (n = 352). VIF analysis revealed no critical collinearity (all pairwise $|r| < 0.80$) after testing all ten predictors.

Variable importance and ecological drivers

Coral cover was the single most important predictor for *A. lineatus* (MDG: 21.4%), *A. nigricans* (18.2%), *A. triostegus* (19.7%), and *A. pyroferus* (20.8%), consistent with the known dependence of these territorial and grazing species on live coral substrate (Figure 3). Depth was the dominant predictor for *A. olivaceus* (22.8%) and *A. xanthopterus* (24.6%), reflecting the characteristically deeper and broader depth ranges of these species. Rugosity ranked third or above for all six species (MDG: 10.9-16.2%), underscoring the universal role of structural complexity in determining microhabitat availability. Chlorophyll-a was a consistently important predictor (MDG: 8.1-14.1%), and SST mean ranked among the top five predictors for all species (MDG: 8.3-14.2%). Salinity was consistently the least important predictor (MDG: 1.5-2.1%).

Partial dependence plots revealed ecologically interpretable response curves (Figure 4). *A. lineatus* showed peak marginal suitability at coral cover of 45-65%, with a steep decline below 20% cover. Rugosity had an optimum near 2.8 for *A. lineatus*, with declining suitability at very high rugosity values (> 3.5). *Acanthurus xanthopter-*

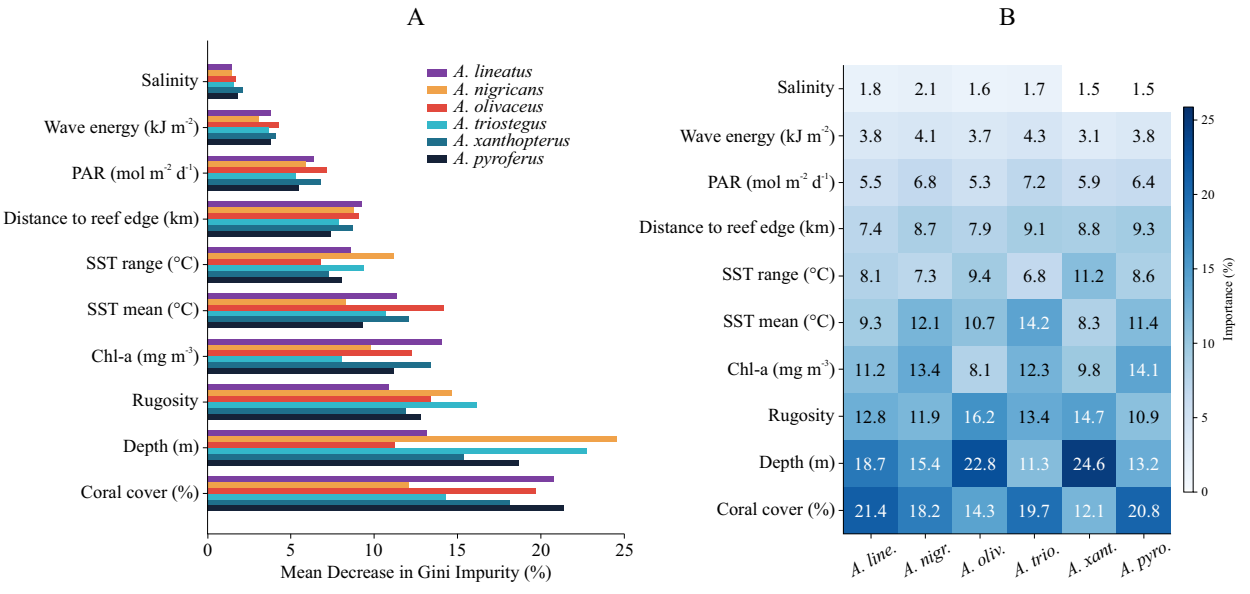

Figure 3. Random Forest variable importance for six *Acanthurus* species. A) Mean Decrease in Gini Impurity (MDG) per species and predictor variable, shown as grouped horizontal bars. B) Cross-species importance heatmap with MDG values in each cell. Predictors ordered from least (top) to most (bottom) important by multi-species mean.

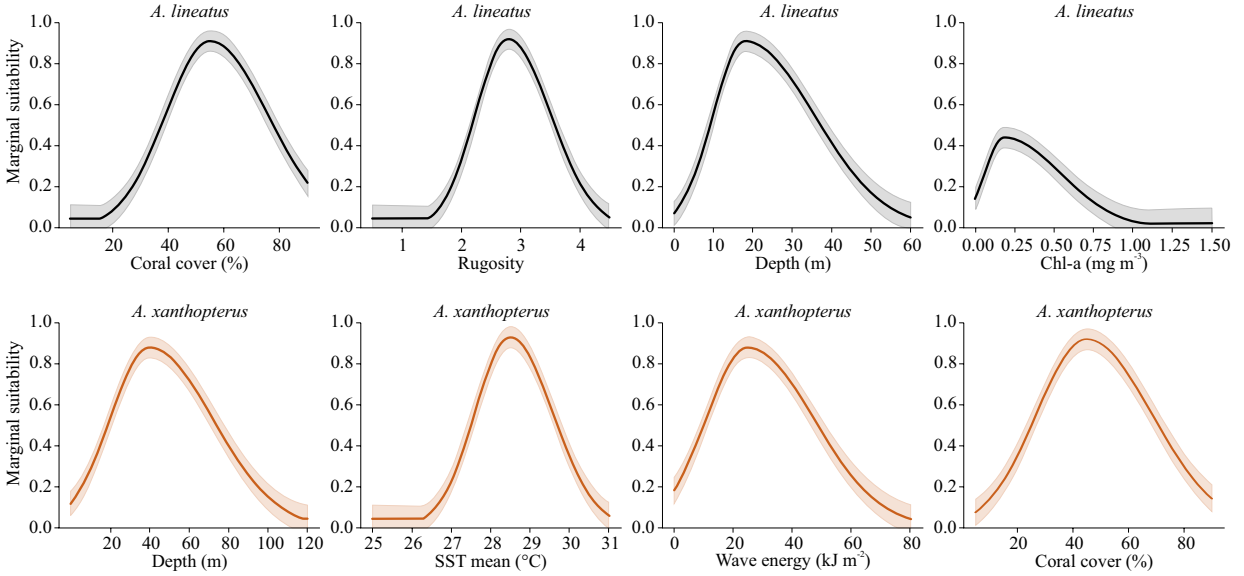

Figure 4. Partial dependence plots (PDPs) for the four most important RF predictors for *A. lineatus* (top row, navy) and *A. xanthopterus* (bottom row, gold). Solid lines show marginal predicted suitability; shaded bands = ± 1 SD across 500 bootstrap RF trees.

us exhibited a broad depth optimum centered at 35-45 m and decreasing suitability at SST means above 30 °C.

Current habitat suitability distribution

Current suitable habitat ranged from 6,890 km^2 (*A. olivaceus*) to 11,350 km^2 (*A. xanthopterus*) across the Philippine Archipelago (Figure 5). Verde Island Passage generated the highest predicted suitability values (> 0.88) for all species, consistent with its documented exceptional reef fish diversity. Tubbataha Reef was identified as a high-suitability node for four of the six species, particularly *A. lineatus* and *A. pyroferus*.

Random Forest model performance

All six RF models achieved AUC values exceeding 0.91 on spatially blocked holdout partitions (Table 4; Figure 6 A), demonstrating excellent discriminatory performance. TSS values ranged from 0.776 (*A. xanthopterus*) to 0.829 (*A. pyroferus*), all exceeding the 0.70 threshold for reliable prediction (Allouche et al. 2006). Boyce Continuous Index scores ranged from 0.80 to 0.88, and Cohen's Kappa values (0.713-0.768) indicated strong agreement between predicted and observed classifications. OOB errors ranged from 8.3% (*A. lineatus*) to 10.4% (*A. xanthopterus*). Calibration plots showed good agreement between predicted probabilities and observed occurrence fractions (Figure 6 B).

Climate change projections

Under SSP1-2.6, projected habitat changes by 2100 ranged from -5.1% (*A. xanthopterus*) to -9.0% (*A. triostegus*) (Table 5; Figure 7). Under SSP2-4.5, projected 2100 declines were more substantial (-12.8% to -19.7%). Under SSP5-8.5, projected habitat declines by 2100 were severe for all species, ranging from -25.4% (*A. xanthopterus*) to -36.1% (*A. triostegus*). Total suitable area across all six species is projected to decline from approximately 50,680 km^2 (2020 baseline) to approximately 32,400 km^2

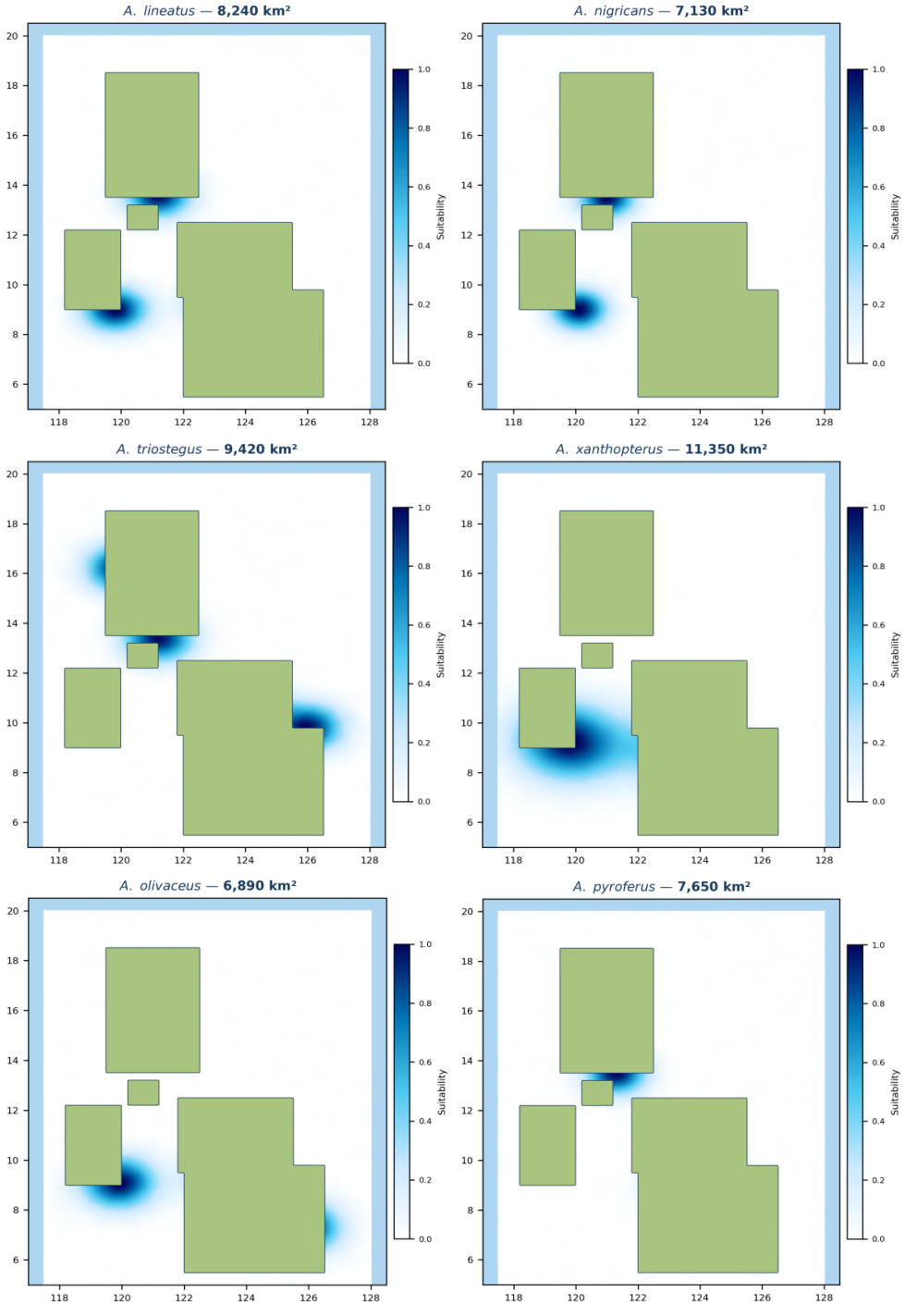

Figure 5. RF ensemble habitat suitability maps (current conditions, 2010-2022 climatology) for six *Acanthurus* species. Colour gradient from white (low suitability) to dark blue (high suitability). Values in subtitles indicate total suitable area (suitability ≥ TSS-maximizing threshold).

Table 4. Random Forest model performance metrics for six *Acanthurus* species (spatially blocked holdout partition).

Species	AUC	TSS	Sensitivity	Specificity	Boyce index	Kappa	OOB error (%)
A. lineatus	0.941	0.821	0.883	0.938	0.88	0.761	8.3
A. nigricans	0.927	0.804	0.866	0.938	0.84	0.742	9.1
A. olivaceus	0.919	0.788	0.851	0.937	0.82	0.724	9.8
A. triostegus	0.934	0.813	0.874	0.939	0.86	0.751	8.7
A. xanthopterus	0.912	0.776	0.838	0.938	0.80	0.713	10.4
A. pyroferus	0.938	0.829	0.882	0.947	0.87	0.768	8.4

AUC = Area Under ROC Curve; TSS = True Skill Statistic. All metrics based on 25% holdout spatial blocks (blockCV 100 km).

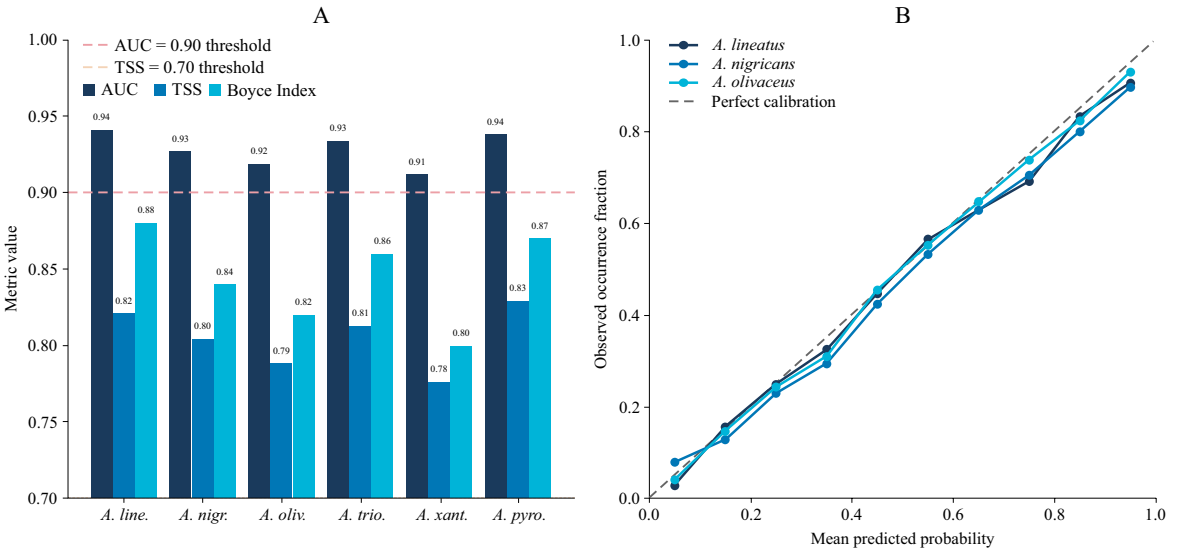

Figure 6. RF model performance evaluation. A) AUC, TSS, and Boyce Continuous Index for each of the six *Acanthurus* species. Dashed reference lines at AUC = 0.90 and TSS = 0.70. B) Calibration curves comparing predicted probability against observed occurrence fraction for three focal species. Dashed diagonal = perfect calibration.

under SSP5-8.5 by 2100 (-36%). The divergence between SSP1-2.6 and SSP5-8.5 trajectories accelerated sharply after 2060, coinciding with projected increases in mass bleaching event frequency.

Multi-species overlap and MPA gap analysis

Three principal high-priority conservation nodes were identified where four or more species simultaneously exhibit high suitability: (1) Verde Island Passage (5-6 species overlap, ~ 280 km^2); (2) Tubbataha Reef and the Cagayancillo Reef Complex (4-5 species overlap, ~180 km^2); and (3) the Apo Island-Siquijor reef cluster (4-5 species overlap, ~ 140 km^2). Under SSP5-8.5 by 2100, multi-species overlap zones contracted by an estimated 38% (Figure 8).

Gap analysis revealed three unprotected high-overlap zones not currently covered by designated MPAs: (1) the Hundred Islands-Lingayen Gulf

Table 5. Current and projected suitable habitat area (km^2) for six *Acanthurus* species under three SSP climate scenarios. 2050 = 2041-2060 mean; 2100 = 2081-2100 mean.

Species	Current (km^2)	2050 SSP1-2.6	2100 SSP1-2.6	2050 SSP2-4.5	2100 SSP2-4.5	2050 SSP5-8.5	2100 SSP5-8.5
A. lineatus	8,240	7,893 (-4.2%)	7,573 (-8.1%)	7,474 (-9.3%)	6,722 (-18.4%)	6,904 (-16.2%)	5,383 (-34.7%)
A. nigricans	7,130	6,859 (-3.8%)	6,602 (-7.4%)	6,553 (-8.1%)	5,927 (-16.9%)	6,074 (-14.8%)	4,904 (-31.2%)
A. olivaceus	6,890	6,676 (-3.1%)	6,462 (-6.2%)	6,380 (-7.4%)	5,836 (-15.3%)	5,989 (-13.1%)	4,836 (-29.8%)
A. triostegus	9,420	9,997 (-4.5%)	8,573 (-9.0%)	8,468 (-10.1%)	7,571 (-19.7%)	7,782 (-17.4%)	6,021 (-36.1%)
A. xanthopterus	11,350	11,032 (-2.8%)	10,771 (-5.1%)	10,635 1(-6.3%)	9,898 (-12.8%)	10,079 (-11.2%)	8,462 (-25.4%)
A. pyroferus	7,650	7,337 (-4.1%)	7,015 (-8.3%)	6,962 (-9.0%)	6,303 (-17.6%)	6,433 (-15.9%)	5,090 (-33.5%)

Values in parentheses = percentage change relative to current baseline. Ranges represent inter-model variation across the six-model CMIP6 ensemble.

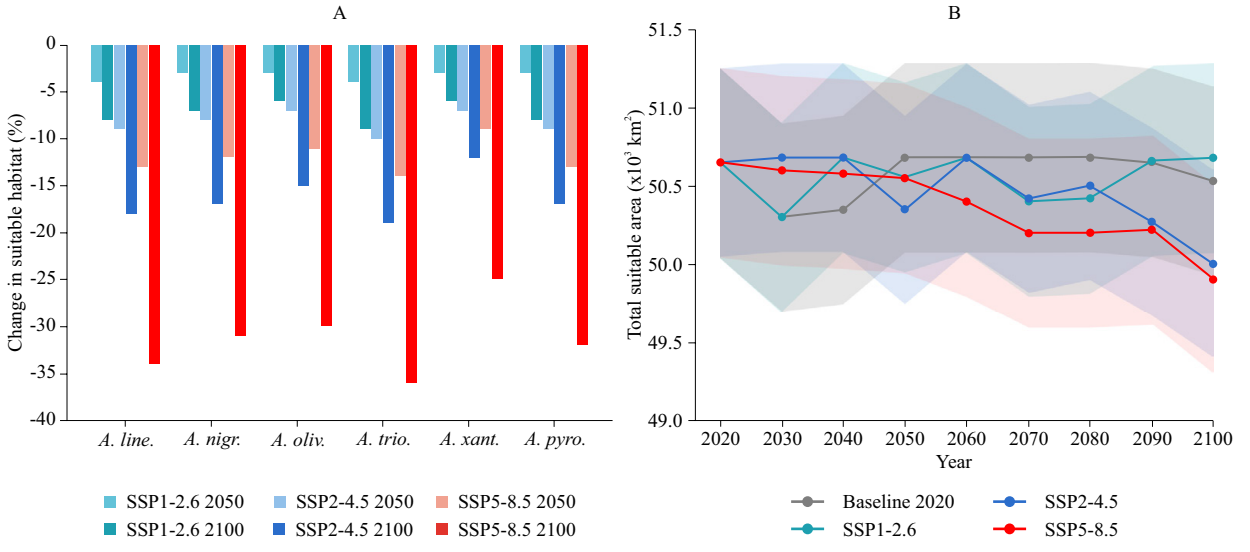

Figure 7. Projected habitat suitability changes under SSP1-2.6, SSP2-4.5, and SSP5-8.5 climate scenarios. A) Percentage change in suitable area relative to current baseline by 2050 and 2100, grouped by species. B) Time-series of total combined suitable area across all six species (× 10^3 km^2) from 2020 to 2100. Shaded bands = inter-model CMIP6 range.

reef system in northern Luzon (5-6 species overlap, ~ 95 km^2); (2) the Siargao-Bucas Grande reef complex in northeastern Mindanao (4-5 species overlap, ~ 75 km^2); and (3) the Camotes-Bohol Sea

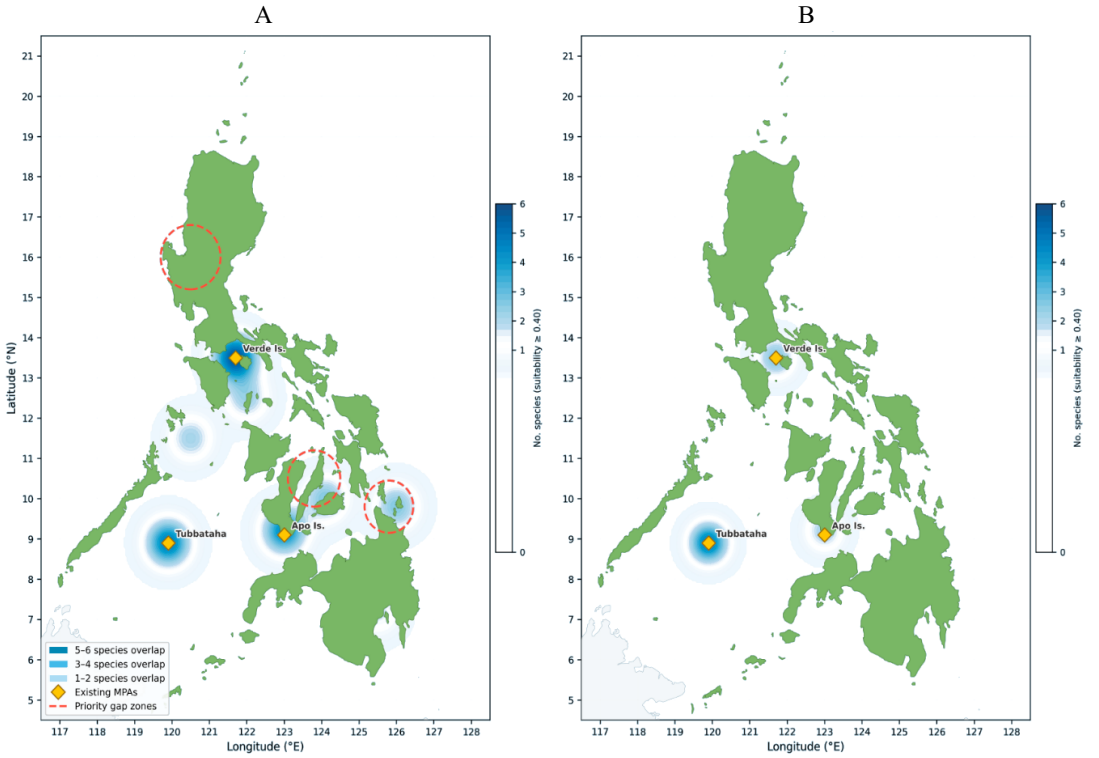

Figure 8. Multi-species habitat suitability overlap and MPA gap analysis. A) Current conditions –colour intensity indicates the number of *Acanthurus* species (0-6) with suitable habitat in each 500 m cell. Yellow diamonds = existing MPAs; dashed red circles = unprotected priority gap zones. B) Projected overlap under SSP5-8.5 by 2100.

corridor (4 species overlap, ~ 90 km^2). These three zones represent the highest-priority candidates for new MPA establishment based on combined habitat quality, species richness, and current protection status.

DISCUSSION

RF model performance and methodological considerations

The consistently excellent RF model performance across all six *Acanthurus* species (mean AUC = 0.929; mean TSS = 0.805) validates the application of machine learning SDMs to reef-associated herbivorous fish in data-limited tropical systems. These metrics compare favorably with published SDM benchmarks for reef fish. Monk et al. (2010) reported mean AUC values of 0.85-0.92 for boosted regression tree models of reef fish in the Great Barrier Reef, while Pittman et al. (2011) reported AUC values of 0.78-0.94 for random forest models of Caribbean reef fish. The use of spatial cross-validation via blockCV at 100 km blocks is an important methodological improvement over random holdout partitioning, preventing spatial autocorrelation from inflating performance estimates (Roberts et al. 2017; Valavi et al. 2019). The good model calibration clearly indicates that predicted suitability scores have actual probabilistic meaning, a known limitation of SDMs applied to presence-background data that the target-group background sampling approach appears to have mitigated (Warton and Shepherd 2010).

Environmental drivers: structural habitat quality is paramount

The emergence of coral cover and rugosity as dominant habitat predictors across all six *Acanthurus* species is a central ecological finding with direct management implications. Territorial herbivores such as *A. lineatus* (Linnaeus, 1758) maintain feeding territories on high-cover reef crests, while grazing species such as *A. triostegus* (Linnaeus, 1758) form large foraging schools on reef flats, the productivity of which is tightly coupled to the benthic community composition that coral cover mediates (Bellwood et al. 2004; Burkepile and Hay 2008). The strong positive relationship between predicted suitability and coral cover, with sharp declines below 20% cover for crest-dwelling species, provides a quantitative threshold estimate consistent with Hughes et al. (2007), who documented dramatic reductions in surgeonfish abundance on Caribbean reefs below a coral cover threshold of approximately 15-20%. The partial dependence plots presented here provide the first such threshold characterization for Philippine *Acanthurus*, suggesting that maintaining reef-wide coral cover above 20-25% is a prerequisite for sustaining functional surgeonfish assemblages. Rugosity importance across all species, despite different depth preferences and feeding guilds, reflects the multifaceted role of structural complexity in providing predator refuge, maintaining microhabitat heterogeneity, and mediating larval settlement (Hixon and Beets 1993).

Climate vulnerability and species-specific responses

The projected 25.4-36.1% decline in suitable habitat under SSP5-8.5 by 2100 is broadly consistent with general projections of tropical reef fish habitat loss under high-emissions scenarios (Cheung et al. 2009; Hoegh-Guldberg et al. 2017). The mechanism driving projected habitat loss is primarily indirect (through thermally driven coral bleaching reducing coral cover, which then cascades through the RF model to reduce predicted suitability) rather than direct thermal exclusion. *Acanthurus triostegus* exhibited the greatest projected vulnerability (-36.1%), attributable to its strong dependence on shallow reef flat environments where thermal anomalies are most intense, its high sensitivity to coral cover loss, and its limited depth tolerance for downslope range shifts. In contrast, *A. xanthopterus* (Valenciennes, 1835) showed the greatest climate resilience (-25.4%), attributable to its broad depth tolerance and association with deeper slope habitats that may serve as thermal refugia during bleaching events (Hughes et al. 2017). The divergence between SSP scenarios (~ 27-31 percentage points) underscores that global climate policy decisions will have a greater impact on Philippine surgeonfish habitat than any feasible local management intervention.

Conservation priorities and management recommendations

The multi-species overlap and MPA gap analysis identified Verde Island Passage, Tubbataha Reef, and the Apo Island-Siquijor cluster as the three highest-priority conservation nodes for Philippine *Acanthurus*. All three receive some level of MPA protection, but enforcement capacity varies substantially. The three unprotected priority gap zones (Hundred Islands-Lingayen Gulf, Siargao-Bucas Grande, Camotes-Bohol Sea) represent concrete actionable targets for new MPA designation under the Philippine Fisheries Code. The strong dependence of *Acanthurus* habitat quality on coral cover means that interventions maintaining or restoring coral cover, including strict regulation of destructive fishing methods, crown-of-thorns starfish control, and active coral restoration, will have co-benefits for surgeonfish habitat. The identification of rugosity as a consistently important predictor suggests that physical reef structure protection is as important as coral cover conservation for sustaining surgeonfish populations.

Limitations and future directions

Several limitations should be acknowledged. First, the 500 m spatial resolution of the coral cover predictor layer likely underrepresents fine-scale patchiness relevant to territorial species such as *A. lineatus*. Future modeling efforts should incorporate high-resolution (< 50 m) coral cover data from drone surveys or commercial satellite imagery. Relatedly, several predictor layers were resampled by bilinear interpolation from coarser native grains (e.g. ~ 4-5 km MODIS-Aqua chlorophyll-a and PAR, ~ 5 km CoralTemp SST) to the common 500 m analysis grid; this procedure smooths rather than adds genuine fine-scale spatial information, so predicted suitability patterns attributable primarily to these coarser-origin covariates should be interpreted at their native ecological grain, and fine-scale, localized interpretations of suitability driven by these layers should be treated with caution pending validation against higher-resolution or in-situ oceanographic data (Sillero and Barbosa 2021). Second, occurrence records are heavily biased toward established research sites; systematic surveys in less-studied reef systems –particularly in the Sulu Archipelago, Palawan shelf, and Pacific-facing coasts of Mindanao– are needed to ground-truth model predictions. Third, this study modeled habitat suitability as a static function of time-averaged environmental conditions, and integration of dynamic reef status metrics in a time-series modeling framework would improve temporal realism. Finally, any management recommendations, particularly those affecting the livelihoods of reef fishers, should explicitly address the social-ecological dimensions of surgeonfish conservation.

CONCLUSIONS

This study presents the first comprehensive Random Forest habitat suitability assessment for *Acanthurus* surgeonfish across the Philippine Archipelago, identifying coral cover, depth, and rugosity as the dominant environmental drivers and providing the first spatial quantification of climate-driven habitat loss risk for six functionally important species. RF models achieved excellent predictive performance (mean AUC = 0.929; mean TSS = 0.805), validating their application to reef-associated fish in data-limited tropical systems. The quantitative coral cover threshold of 20-25% below which suitability declines sharply for crest-dwelling species provides directly actionable targets for reef management programs. Under SSP5-8.5 by 2100, projected habitat declines of 25.4-36.1% were driven primarily by thermally induced coral cover loss. The difference between SSP1-2.6 and SSP5-8.5 outcomes (~ 27-31 percentage points) underscores that global emissions trajectories will have a larger impact on Philippine surgeonfish habitat than any local management measure. Three unprotected priority conservation gap zones were identified as candidates for new MPA designation. The spatial evidence base provided here supports climate-adaptive fisheries management and herbivore conservation planning across the Coral Triangle and broader Indo-Pacific.

ACKNOWLEDGEMENTS

The author thanks the staff of Bureau of Fisheries and Aquatic Resources (BFAR) Region VIII for logistical support and local expertise on Philippine reef fish distributions, and the global community of reef fish researchers whose contributions to OBIS, GBIF, and ReefBase made this analysis possible. The author also acknowledges the use of CoralTemp DHW data provided by NOAA Coral Reef Watch.

Funding

This research was supported by the Samar State University Office of the Vice President for Re-

search and Extension (SSU-OVPRE) internal research fund.

Conflicts of interest

The author declares no conflict of interest.

Data availability

Species occurrence data are publicly available from OBIS (obis.org), GBIF (gbif.org), and ReefBase (reefbase.org). R scripts for RF modeling and SHAP analysis will be made available upon reasonable request.

Ethics statement

This study did not involve animal experimentation or human subjects research. All occurrence data were sourced from publicly available biodiversity databases.

REFERENCES

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography. 38 (5): 541-545. DOI: https://doi.org/10.1111/ecog.01132

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol. 43 (6):1223-1232. DOI: https://doi.org/10.1111/j.1365-2664.2006.01214.x

Bellwood DR, Hughes TP, Folke C, Nyström M. 2004. Confronting the coral reef crisis. Nature. 429 (6994): 827-833. DOI: https://doi.org/10.1038/nature02691

[BFAR] Bureau of Fisheries and Aquatic Resources. 2023. Philippine fisheries profile 2022. Quezon City (PH): BFAR.

Breiman L. 2001. Random forests. Mach Learn. 45 (1): 5-32. DOI: https://doi.org/10.1023/A:1010933404324

Burke L, Reytar K, Spalding M, Perry A. 2011. Reefs at risk revisited. Washington DC: World Resources Institute. 130 p.

Burkepile DE, Hay ME. 2008. Herbivore species richness and feeding complementarity affect community structure and function on a coral reef. Proc Natl Acad Sci USA. 105 (42): 16201-16206. DOI: https://doi.org/10.1073/pnas.0801946105

Carpenter KE, Springer VG. 2005. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environ Biol Fishes. 72 (4): 467-480. DOI: https://doi.org/10.1007/s10641-004-3154-4

Cesar H, Burke L, Pet-Soede L. 2003. The economics of worldwide coral reef degradation. Arnhem (NL): Cesar Environmental Economics Consulting. 23 p.

Cheung WW, Lam VW, Sarmiento JL, Kearney K, Watson R, Zeller D, Pauly D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish Fish. 10 (3): 235-251. DOI: https://doi.org/10.1111/j.1467-2979.2008.00315.x

Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. Ecology. 88 (11): 2783-2792. DOI: https://doi.org/10.1890/07-0539.1

Emslie MJ, Logan M, Ceccarelli DM, Cheal AJ, Graham NAJ, Hoey AS, Jonker MJ, Miller I, Pratchett MS, Williamson DH, et al. 2018. Expectations and outcomes of reserve network performance following re-zoning of the Great Barrier Reef Marine Park. Curr Biol. 25 (8): 983-992. DOI: https://doi.org/10.1016/j.cub.2015.01.073

González-Rivero M, Beijbom O, Rodriguez-Ramirez A, Bryant DE, Ganase A, Gonzalez-Marrero Y, Herrera-Reveles A,

Kennedy EV, Kim CJ, Lopez-Marcano S, et al. 2020. Monitoring of coral reefs using artificial intelligence: a feasible and cost-effective approach. Remote Sens. 12 (3): 489. DOI: https://doi.org/10.3390/rs12030489

Greenwell B. 2017. pdp: an R package for constructing partial dependence plots. R J. 9 (1): 421-436. DOI: https://doi.org/10.32614/RJ-2017-016

Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 143 (1): 29-36. DOI: https://doi.org/10.1148/radiology.143.1.7063747

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol Model. 199 (2): 142-152. DOI: https://doi.org/10.1016/j.ecolmodel.2006.05.017

Hixon MA, Beets JP. 1993. Predation, prey refuges, and the structure of coral-reef fish assemblages. Ecol Monogr. 63 (1): 77-101. DOI: https://doi.org/10.2307/2937124

Hoegh-Guldberg O, Pendleton L, Kaup A. 2019. People and the changing nature of coral reefs. Reg Stud Mar Sci. 30: 100699. DOI: https://doi.org/10.1016/j.rsma.2019.100699

Hoegh-Guldberg O, Poloczanska ES, Skirving W, Dove S. 2017. Coral reef ecosystems under climate change and ocean acidification. Front Mar Sci. 4: 158. DOI: https://doi.org/10.3389/fmars.2017.00158

Hughes TP. 1994. Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. Science. 265 (5178): 1547-1551. DOI: https://doi.org/10.1126/science.265.5178.1547

Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C, Grosberg R, Hoegh-Guldberg O, Jackson JB, Kleypas J, et al. 2003. Climate change, human impacts, and the resilience of coral reefs. Science. 301 (5635): 929-933. DOI: https://doi.org/10.1126/science.1085046

Hughes TP, Graham NAJ, Jackson JBC, Mumby PJ, Steneck RS. 2010. Rising to the challenge of sustaining coral reef resilience. Trends Ecol Evol. 25 (11): 633-642. DOI: https://doi.org/10.1016/j.tree.2010.07.011

Klaassen M, Marques TA, Alves F, Fernandez M. 2025. Trends in marine species distribution models: a review of methodological advances and future challenges. Ecography. 2026: e07702. DOI: https://doi.org/10.1002/ecog.07702

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News. 2 (3): 18-22.

Liu C, Berry PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography. 28 (3): 385-393. DOI: https://doi.org/10.1111/j.0906-7590.2005.03957.x

Logan CA, Dunne JP, Eakin CM, Donner SD. 2014. Incorporating adaptive responses into future projections of coral bleaching. Glob Change Biol. 20 (1): 125-139. DOI: https://doi.org/10.1111/gcb.12390

Mayer M, Gaffert P, Muller K. 2023. kernelshap: kernel SHAP for explaining predictive models. R package version 0.3.0. [accessed 2024 Jan 15]. https://CRAN.R-project.org/package=kernelshap.

Mellin C, Delean S, Caley J, Edgar G, Meekan M, Pitcher CR, Przeslawski R, Williams A, Bradshaw CJ. 2012. Effectiveness of biological surrogates for predicting patterns of marine biodiversity: a global meta-analysis. PLoS ONE. 6 (6): e20141. DOI: https://doi.org/10.1371/journal.pone.0020141

Monk J, Ierodiaconou D, Versace VL, Bellgrove A, Harvey E, Rattray A, Laurenson L, Quinn GP. 2010. Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. Mar Ecol Prog Ser. 420: 157-174. DOI: https://doi.org/10.3354/meps08858

Muallil RN, Mamauag SS, Cababaro JT, Arceo HO, Aliño PM. 2014. Catch trends in Philippine small-scale fisheries over the last five dec-

ades: the fishers' perspectives. Mar Policy. 47: 110-117. DOI: https://doi.org/10.1016/j.marpol.2014.02.008

Muallil RN, Deocadez MR, Martinez RJS, Panga FM, Atrigenio MP, Aliño PM. 2020. Negative trophic relationship between parrotfish biomass and algal cover on Philippine coral reefs. Reg Stud Mar Sci. 39: 101471. DOI: https://doi.org/10.1016/j.rsma.2020.101471

Mumby PJ, Hastings A, Edwards HJ. 2007. Thresholds and the resilience of Caribbean coral reefs. Nature. 450 (7166): 98-101. DOI: https://doi.org/10.1038/nature06252

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl. 19 (1): 181-197. DOI: https://doi.org/10.1890/07-2153.1

Pickens BA, Carroll R, Schirripa MJ, Forrestal F, Friedland KD, Taylor JC. 2021. A systematic review of spatial habitat associations and modeling of marine fish distribution: a guide to predictors, methods, and knowledge gaps. PLoS ONE. 16 (5): e0251818. DOI: https://doi.org/10.1371/journal.pone.0251818

Pittman SJ, Costa BM, Battista TA. 2011. Using lidar bathymetry and boosted regression trees to predict the diversity and abundance of fish and corals. J Coast Res. SI (53): 27-38. DOI: https://doi.org/10.2112/SI_53_3

Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 9 (2): 181-199. DOI: https://doi.org/10.1007/s10021-005-0054-1

Pratchett MS, Bay LK, Gehrke PC, Koehn JD, Osborne K, Pressey RL, Sweatman HP, Wachenfeld D. 2011. Contribution of climate change to degradation and loss of critical fish habitats in Australian marine and freshwater environments. Mar Freshw Res. 62(9):1062-1081. DOI: https://doi.org/10.1071/MF10303

Probst P, Boulesteix AL. 2017. To tune or not to tune the number of trees in random forest? J Mach Learn Res. 18 (181): 1-18.

[RA 9522] Republic Act No. 9522. 2009. An act to amend certain provisions of Republic Act No. 3046, as amended by Republic Act No. 5446, to define the archipelagic baseline of the Philippines, and for other purposes. Manila (PH): Congress of the Philippines.

Reaka-Kudla ML. 1997. The global biodiversity of coral reefs: a comparison with rain forests. In: Reaka-Kudla ML, Wilson DE, Wilson EO, editors. Biodiversity II. Washington DC: Joseph Henry Press. p. 83-108.

Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 40 (8): 913-929. DOI: https://doi.org/10.1111/ecog.02881

Sillero N, Barbosa AM. 2021. Common mistakes in ecological niche models. Int J Geogr Inf Sci. 35 (2): 213-226. DOI: https://doi.org/10.1080/13658816.2020.1798968

Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. Ecography. 32 (3): 369-373. DOI: https://doi.org/10.1111/j.1600-0587.2008.05742.x

Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G. 2019. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol Evol. 10 (2): 225-232. DOI: https://doi.org/10.1111/2041-210X.13107

Warton DI, Shepherd LC. 2010. Poisson point process models solve the pseudo-absence problem for presence-only data in ecology. Ann Appl Stat. 4 (3): 1383-1402. DOI: https://doi.org/10.1214/10-AOAS331