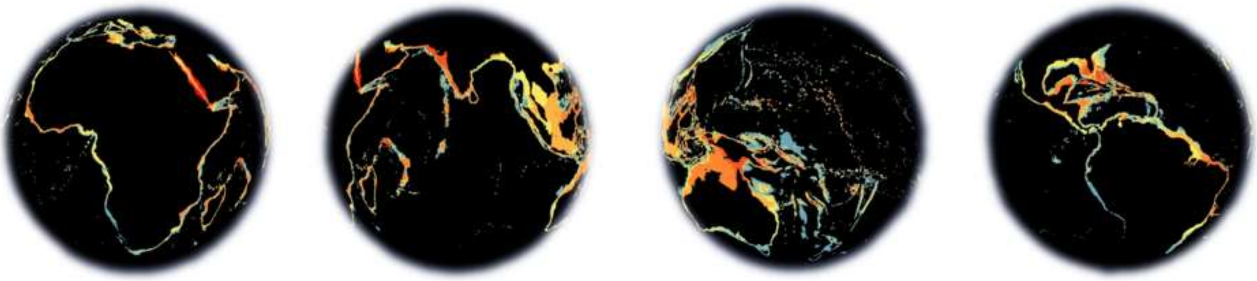

Predicting the global distribution of tropical coral reefs using machine learning algorithms

Vitalis Dubininkas[1,2] & Stanley Mastrantonis[1,3]

January 2017

---

[1] Oshana Environmental Consulting; 6 Empen Way; Hillarys, WA 6025; Australia
[2] vitalis@oshana.net
[3] stan@oshana.net

Abstract

Recent events of widespread coral bleaching allude to the significant impact of climate change and anthropogenic pressure upon the shallow-water marine ecosystems known as coral reefs. Given the limitations of underwater satellite imagery and predictions of average sea surface temperature rise to be 2°C by 2060, novel methods that aid conservation on a large scale are required. Previous studies have attempted to map the global distribution of tropical corals by directly calculating the area of reefs using nautical charts and by calculating continental shelf area using predictive techniques. Conversely, the present study aimed to predict the spatial distribution of tropical coral reefs by incorporating environmental data and elements of machine learning into three different species distribution models. By using habitat suitability of hermatypic corals as a proxy for reef presence, this study was able to predict the global distribution of tropical coral reefs with a 20 x 20 km resolution. It was observed that bathymetry, sea-surface temperatures, salinity, and pH significantly affected the likelihood of observing coral reef presence in all three models. The models also performed favourably under sensitivity analysis (AUC > 0.84), especially when compared to empirical observations of known coral reef systems. Areas which were predicted to have marginal-to-high suitability were used to approximate the spatial coverage on a global scale. Using the combined results of the models, it was estimated that suitable habitat for hermatypic corals is approximately 3.53 ± 0.18% of the world's oceans. Nevertheless, calculations of pristine habitat (i.e. $p \geq 0.9$ of finding a coral reef) were noted to be closer to 1.0 % of the world's oceans. The outputs of the models, as well as the data that was generated as a result of this study, should be of interest to coral ecologists, wildlife biologists, fisheries managers, and groups within the petroleum industry.

Introduction

Coral reefs are diverse ecosystems that are characterized by the presence of organic calcium carbonate deposits (Milliman, 1974). Though many biotic and abiotic components interact synergistically to form a coral reef (Carleton Ray, 1996; Polónia *et al*., 2015), one taxon has been observed to have a disproportionally large effect on the ecosystem (Kerry & Bellwood, 2015). It is argued that hermatypic (Scleractinian) corals are keystone species in these ecosystems, as they directly create habitat complexity (Alvarez-Filip *et al*., 2013); which provides shelter for a range of marine organisms, such as teleosts (Chabanet *et al*., 1997) and crustaceans (Lozano-Álvarez *et al*., 2007). In addition to habitat complexity, the soft tissue of hermatypic corals also act as a source of nourishment for many species of teleosts and invertebrates; such as members of the Scaridae (parrotfish; Bonaldo *et al*., 2012), Chaetodontidae (butterfly fish; Cole *et al*., 2011), Tetraodontidae (pufferfish; Palacios *et al*., 2014), Amphinomidae (fireworms; Wolf & Nugues, 2013), Diadematidae (long-spine sea urchins; Bak & van Eys, 1975), and Acanthasteridae (venomous starfish; Walbran *et al.*, 1989) families. The presence of hermatypic corals also indirectly affects other marine organisms, such as pinnipeds (Parrish *et al*., 2000) and cheloniids (León & Bjorndal, 2002), by providing them a sheltered feeding ground. Furthermore, hermatypic corals are also known to affect the economy of coastal communities in many different ways (Moberg & Folke, 1999); such as acting as a shelter for juvenile specimens of commercial fishes (McClanahan, 1994), being a tourism attraction (Brander *et al.*, 2007), and by acting as a buffer to storm surge (Stoddart, 1971). Given the importance of hermatypic corals, it is important to know the spatial extent of suitable habitat for these corals, and the factors which influence it.

Previous studies have identified a number of natural, anthropogenic, or a combination of these factors, to be responsible for the regional decline or absence of living corals (Birkeland, 2004). In terms of natural factors, coral assemblages have been shown to be negatively affected by excessive ultra-violet radiation (Aranda *et al.*, 2011), abnormally high water temperatures (Lough &

van Oppen, 2009), storms (Foster *et al*., 2011; Osborne *et al*., 2011), earthquakes (Aronson *et al*., 2012), and microbial infections (Aeby *et al.*, 2011). Corals can also be negatively affected by sedimentation, which is a result of natural and/or urban runoff (Fabricius, 2005). Other anthropogenic factors with detrimental effects to coral assemblages include trampling by tourists (Meyer & Holland, 2008), marine pollution (Dubinsky & Stambler, 1996), dredging (Bak, 1978), and other physical damage induced by aquatic vessels (Davis, 1977). Nevertheless, given the fact that coral reefs have existed for thousands of years (Caley & Richards, 1956; Jackson, 1992), it can be inferred that natural environmental conditions are the factors that primarily limit the spatial distribution of hermatypic corals.

In terms of these environmental parameters, bathymetry is one of the elements that has been observed to be limiting the presence of corals (Mumby *et al*., 2004). Since light attenuation increases with depth and turbidity (Markager & Vincent, 2000), most hermatypic coral species are confined to shallow water (Fricke & Schuhmacher, 1983; Grigg, 2006). In addition to bathymetry, ambient water temperature is another element that has been observed to have controlling effects on tropical coral distribution (Coles, 1976). Since corals and zooxanthellae are ectotherms, their metabolic rates are governed by ambient water temperatures; which in turn, are governed by water currents and insolation (Miller & Wheeler, 2012). Similar to most ectotherms, hermatypic corals can also survive in a relatively wide range of temperatures (Macintyre & Pilkey, 1969; Jokiel & Coles, 1977); but they seem to perform at an optimum in a niche of temperatures (Coles & Jokiel, 1977). Other elements which have also been observed to influence the spatial distribution of corals are salinity and pH. Salinity has been observed to affect the metabolic rates of corals and zooxanthellae (Muthiga & Szmant, 1987); whereas pH has been observed to directly affect the calcification rates of corals (Marubini & Atkinson, 1999) and the metabolic rate of zooxanthellae (Kühl *et al*., 1995). Overall, hermatypic corals exist in a narrow range of bathymetry, temperature, salinity, and pH values.

The four aforementioned elements are considered fundamental, because they directly affect habitat suitability for hermatypic corals. If any of these environmental parameters was unsuitable, then it is highly improbable that living corals would be present; which in turn would affect the likelihood of finding a tropical coral reef there. Though studies have been conducted which empirically surveyed regions of the ocean for coral reefs, a lack of funding and resources has hindered a fine-scale, global empirical study (Mumby *et al*., 1997). The purpose of this study is to create a model, which synthesizes empirical data, to predict the spatial extent and distribution of tropical coral reefs; based on the fundamental niche of hermatypic corals. Similar models have previously been constructed for deep-water coral (Bryan & Metaxas, 2007; Davies *et al.*, 2008) and invasive macroalgae (Tyberghein *et al*., 2012) taxa; which have yielded promising results. Our hypothesis is that the known distribution of tropical coral reefs will be successfully predicted by values of bathymetry, temperature, salinity, and pH.

Methods

In order to predict suitable coral habitat for the planet, geographical variables that effectively covered the oceans were required. Having determined the primary factors that directly and indirectly affect coral survivorship (i.e. temperature, depth, pH and salinity), several global databases were interrogated and evaluated for use in distribution models. Oceanographic variables for sea-surface temperature, pH, and salinity were sourced from the World Ocean Database (NOAA, 2013). Specifically, the raw data from the Ocean Station Dataset were selected and extracted, using Visual Studio, and imported into geographic information system software (*ArcMap*; McCoy *et al*., 2001). Point data for each of the environmental parameters, for every station (temperature: *N* = 611,664; pH: *N* = 492,112; salinity: *N* = 611,531), were interpolated into continuous surfaces using an inverse distance weighting (IDW) algorithm (Childs, 2004). To minimise error, power parameters (i.e. the influence from increasingly distant neighbours) for the IDW algorithm were optimised using cross-validation techniques. Depth data was sourced from the General Bathymetric Chart of the Ocean (GEBCO; Fisher *et al.*, 1992). Six 120° × 80°, 30-second arc interval grid ASCII files were individually extracted, and combined into a single mosaic RASTER dataset; effectively representing bathymetry values for the entire globe. Using shapefiles of the world's land masses, all terrestrial cells were removed from the RASTER; resulting in a bathymetric layer with a cell size of 0.0083°× 0.0083°. Geotagged data of known coral reefs was sourced from an online database (ReefBase, 2016). These observations formed the basis of presence values within the distribution model (N = 5,140); while pseudo-absences (Chefaoui, 2008; Mateo *et al.*, 2010; Barbet-Massin *et al.*, 2012) were derived on the basis of 40 – 200 km radial buffers around each of the presence observations (Figure 1). The presence and pseudo-absence points were laid over the environmental parameter layers, which formed a set of binomial response values. Subsequently, the values of each of the environmental parameters were extracted at every intersection with a presence or absence point; which formed a set of predictor variables.

A prediction layer (WGS 1984 World Mercator) was created for the entire ocean that included the temperature, pH, salinity, and depth parameters; all of which were resampled into uniform geographic extents and cell sizes. Grid-squares of 20 km^2 (N = 2,354,392) were generated, and overlaid the prediction layer. At each individual grid square, the corresponding values for the environmental conditions were extracted, and imported into statistical software (R Development Core Team, 2016) for predictive analysis (Wood, 2001). To predict the spatial extent of suitable habitat for hermatypic corals, the presence, absence, and environmental parameter data were incorporated into several species distribution models (Austin, 2002; Guisan & Thuiller, 2005; Elith & Leathwick, 2009). A generalised linear model (GLM) was fit using a logistic link function (Zuur, 2007), which took the following form:

$$logit(\pi) = log\frac{\pi}{(1-\pi)} = \beta o + temperature + salinity + pH + bathymetry\ ;$$

where π represents the probability of presence, and βo represents the intercept. A generalised additive model (GAM) with a logistic link was also fit, which took the following form:

$$f(x) = \beta o + f(temperature) + f(salinity) + f(pH) + f(bathymetry)\ ;$$

where βo represents the additive model intercept, and f represents the smoothed functions of explanatory variables (Zuur, 2007). Knots were penalised within the model, using cross-validation to optimise the smooth functions and minimise residual error. Cross-validation was set to 10 k-folds for comparison, and shrinkage was also applied to model all model terms. Nevertheless, no terms were removed due to parameter shrinkage. In both the GAM and the GLM, overdispersion was also tested for by fitting the model using a quasibinomial distribution and assessing the dispersion parameters. Additionally, a classification and regression tree (CRT) was fit to the data, which took the following form:

$$f(x) = \sum_{m=1}^{M} c_m \cdot 1_{(X \in R_m)}\ ;$$

where *X* is a matrix of predictor values, *R* represents regions of predictor space, and *c* represents the penalty term against divergent nodes (James, 2013). Classification tree regression incorporated a binary recursive splitting algorithm, to separate these regions within the predictor space. In order to minimise the residual error, cross-validation (i.e. using half the data as a training set and the other half as a test set) was conducted to fit a tree with an optimal number of region splits.

The outputs of all three models were used to predict the probability of hermatypic coral presence in every grid square. The resulting probabilities were interpreted to be the spatial extent of suitable habitat for hermatypic corals; where 0.99 would indicate ideal environmental conditions for a coral reef, and 0 would indicate an area with naturally inhospitable environmental conditions. Model predictions were compared to empirical data, and qualitatively assessed on ecological and oceanographic principles. In addition to qualitative validation, all model outputs were also validated using receiver-operating characteristics (ROC; Leathwick *et al*, 2006; Ficetola, 2007), and by deriving the cumulative area under the curve (AUC); where an AUC of 1.0 would indicate a model that flawless fit the input the data, and an AUC of 0.50 would indicate model predictions that are no better than random chance. All cells within the study area, which achieved a predicted probability of ≥ 0.50 for hermatypic coral presence, were used to calculate the spatial extent of suitable habitat for coral reefs. Since CRT predictions are binary, a cut off of $p = 0.50$ was chosen; which allowed for a more appropriate comparison of the CRT to the linear models. Calculations of the spatial extent were performed using NOAA's ETOPO1 global relief model, which provided estimates of oceanic basin surface area. The outputs of the three models were combined, to yield a mean and standard deviation of the predicted spatial extent of tropical coral reef coverage.

Results

All terms within the GLM were observed to be significant at the $\alpha = 0.05$ level ($R^2 = 0.26$; df = 10,235); with null and residual deviances being calculated to be 14,195 and 10,465, respectively. Similar to the GLM, all of the smooth terms in the GAM were also observed to be significant at the $\alpha = 0.05$ level ($R^2 = 0.54$; df = 10,231). Assessment of diagnostic plots for both the GLM and GAM indicated that residuals satisfied the assumption of normality, when modelled using a log link. Nevertheless, the GAM performed much better in this respect, when compared the GLM. In order to test if residuals may exhibit issues of spatial autocorrelation, a generalised least square component was further incorporated into the models, but it yielded no improvements in model fit. In regards to the CRT, cross-validation between the training and test sets resulted in a root-mean-squared error of 1.37. The CRT, itself, consisted of 6 nodes with an associated mean deviance of 0.07.

The multivariate effects of all three models were also investigated (Figure 2), which showed the niche of environmental conditions in which hermatypic corals are likely to be found (Figure 2 (b)). In terms of probabilities being output from the models, large-scale geographic comparisons showed that predictions of all three models varied slightly. Yet, the spatial extents of all of the models were largely in congruence (Figure *3*). Model predictions were also observed to have a significant amount of spatial overlap with previous empirical observations of coral reefs (e.g. ReefBase, UNEP; Figure 4). Model validation using ROC showed that the CRT was the most precise in fitting the empirical observations of coral reefs, with an AUC of 0.96. In comparison, the GLM and GAM were observed to have AUC scores of 0.84 and 0.94. respectively (Figure 5). In terms of the predicted spatial extent of suitable habitat for hermatypic corals, all three models estimated a relatively similar area. Given a probability cut-off point of 0.5, the GLM and GAM predicted approximately 3.65% and 3.62% of the ocean to be suitable habitat, respectively. In turn, the CRT predicted suitable habitat to be 3.32%. From the combined results of the models, it was estimated that suitable habitat for hermatypic corals is approximately 12,766,179 ± 659,946 km^2 of the world's oceans.

Discussion

Two broad approaches have previously been used to estimate the spatial extent of tropical coral reefs: calculating reef shelf area using predictive techniques (Milliman, 1974; Smith, 1978; Copper, 1994; Couce, *et al.*, 2012) and direct calculation of reef area using nautical charts (Newell, 1971; De Vooys, 1979). The present study aimed to create a model, which synthesized empirical data, to predict the spatial extent and distribution of tropical coral reefs; based on the fundamental niche of hermatypic corals. This fundamental niche was defined as a range of bathymetry, sea-surface temperatures, salinity, and pH values which determines the presence and absence of coral reefs. Using a GLM, GAM and CRT, the probability of observing a tropical coral reef was predicted into 2,354,392 grid squares, which had an individual area of 20 km^2. In general, it was observed that the spatial extent of tropical coral habitat predicted by all three models was in corroboration with previous remote sensing (NASA, 2007) and data compiling efforts (UNEP-WCMC, 2010; ReefBase, 2016) (Figure 4). Using AUC values of ROC curves as a form of model validation, it was noted that all of the models were performing favourably, with model predictions being significantly better than random chance. It was also noted that the CRT and GAM were more sensitive in error detection than the GLM. This is likely due to the flexibility of the smooth functions in the GAM, and the recursive partitioning incorporated in the CRT; that allowed a more accurate detection of false-positive and true-negative observations, which would have diminished the overall AUC score of the models. Nonetheless, since all AUC values exceeded the threshold for model rejection (AUC = 0.50), it was assumed that all model predictions accurately detected areas that are likely to exhibit an absence or presence of coral reefs.

The outputs of these models are unique, as they can be used to explain why coral reefs are likely to be present in certain areas, and absent in others. For example, coastal areas of the Black Sea were observed to have suitable sea surface temperatures, pH, and bathymetry values for hermatypic coral presence. However, the low salinity in this region was observed to be responsible for the low predicted probability of coral reef presence. Similar observations were also made for the continental shelf off of the coast of Cameroon and Nigeria; as well as most of the continental shelf found in the Gulf of Mexico. Here, it was observed that areas with suitable pH, bathymetry, and sea surface temperatures exist. Yet, the suitability of the habitat was predicted to be relatively low due to the naturally brackish conditions. In terms of highly suitable habitat, it was observed that all three models favoured the coastline of Eritrea, Yemen, and western Saudi Arabia. Naturally arid conditions, combined with optimal depth, pH, and thermal conditions led to model predictions favouring this region. Interestingly, it was also observed that relatively sharp gradients in probabilities exist in the areas that are predicted to have coral reef presence; which supports Hughes *et al.* (2012) findings of complex latitudinal gradients in coral community composition, and Barkley *et al.* (2015) observations of coral community composition shifts across a natural pH gradient. It can be inferred that the gradients predicted by the models are a result of either the singular or combined effects of thermohaline circulation and topography, which influences habitat suitability by directly affecting the values of pH, salinity, sea-surface temperature, and bathymetry.

Remarkably, it was also noted that empirical observations of coral reefs were located in areas which had predicted low probabilities of coral reef presence ($p \leq 0.3$). This is an important observation because it highlights the opportunistic nature of hermatypic corals, and the ability of certain species to survive in areas with relatively inhospitable environmental conditions (Macintyre & Pilkey, 1969; Spalding, *et al.*, 2001).

The outputs of the three models were also used to predict the spatial extent of suitable habitat for tropical coral reefs. It was observed that all three models predicted a relatively similar area, which was calculated to be approximately 3.53 ± 0.18% of the world's oceans. The resulting estimate is slightly larger than previous approximations of the spatial extent of coral reefs (Newell, 1971; Smith, 1978; Spalding & Grenfell, 1997; Milliman, 1974; Copper, 1994). The discrepancies in the estimated area are likely a result of using a probability cut off of 0.50. That is, it is an arbitrarily picked value, which governed the computer's decision to either deem an area as suitable habitat or not. It was observed that if the cut off was set to a higher value, then the spatial extent of suitable habitat gradually diminished. Conversely, if the cut off was set to a smaller value, then the extent of suitable habitat increased. Though the models were relatively precise in predicting known tropical coral reef habitat, they also had their limitations. Namely, there is an unknown degree of error associated with the environmental parameter and empirical coral reef observation data sets; which could not be accounted for in the present study. Once more global data sets of environmental parameters become available, and the computational capacity of personal computers increases, then a more comprehensive model can be made. Ideally, this model should incorporate the four factors that compose the fundamental niche of hermatypic corals, as well as all of the factors that are known to affect the metabolism and survivorship of corals and zooxanthellae, such as: substrate availability (Bowden-Kerby, 2001), relative concentration of dissolved organic matter (Kleypas *et al.*, 1999), solar radiation (Jokiel & York, 1982), wave energy (Bradbury & Young, 1981), carbonate/aragonite saturation state (Kleypas *et al.*, 1999), and other factors (Davis, 1977; Bak, 1978; Dubinsky & Stambler, 1996; Fabricius, 2005; Meyer & Holland, 2008; Lough & van Oppen, 2009; Aeby *et al*., 2011; Aranda *et al*., 2011; Foster *et al*., 2011; Osborne *et al*., 2011; Aronson *et al*., 2012). At that point, it would be wise to evaluate the trade-off between adding more explanatory variables to the model and losing model parsimony. In general, the spatial coverage of tropical coral reefs should be of interest in future studies. This is due to the fact that previous studies have reported results that are largely different from one another (Newell, 1971; Milliman, 1974; Smith, 1978; Copper, 1994; Spalding & Grenfell, 1997), and the results of the present study have yielded estimates that are slightly more liberal than previous approximations (Goodman, *et al.*, 2013).

Acknowledgements

The authors would like to thank NOAA and GEBCO for proving global data of environmental parameters, and ReefBase for providing the locations of known coral reefs. The authors would also like to thank Pia Ricca for proofreading the manuscript. Vitalis Dubininkas would like to thank his wife for motivation throughout this project. Stanley Mastrantonis would like to thank his family. The authors would like to thank the MSc/MRes Applied Marine & Fisheries Ecology programme staff at the University of Aberdeen.

Conflicts of Interest

There are no competing financial interests.

References

Aeby, G. S., Williams, G. J., Franklin, E. C., Kenyon, J., Cox, E. F., Coles, S. & Work, T. M. (2011) Patterns of coral disease across the Hawaiian archipelago: relating disease to environment. *PLOS ONE*, **6**(5): e20370, DOI: 10.1371/journal.pone.0020370

Alvarez-Filip, L., Carricart-Ganivet, J. P., Horta-Puga, G. & Iglesias-Prieto, R. (2013) Shifts in coral-assemblage composition do not ensure persistence of reef functionality. *Scientific Reports*, **3**, DOI: 10.1038/srep03486

Aranda, M., Banaszak, A. T., Bayer, T., Luyten, J. R., Medina, M. & Voolstra, C. R. (2011) Differential sensitivity of coral larvae to natural levels of ultraviolet radiation during the onset of larval competence. *Molecular Ecology,* **20**(14), 2955-2972.

Aronson, R. B., Precht, W. F., Macintyre, I. G. & Toth, L. T. (2012) Catastrophe and the life span of coral reefs. *Ecology*, **93**, 303-313.

Austin, M. P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157,** 101–118.

Bak, R. P. M. (1978) Lethal and sublethal effects of dredging on reef corals. *Marine Pollution Bulletin,* **9**, 14-16.

Bak, R. P. M. & van Eys, G. (1975) Predation of the sea urchin Diadema antillarum Philippi on living coral. *Oecologia*, **20**, 111-115.

Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3,** 327–338.

Barkley, H. C., Cohen, A. N., Golbuu, Y., Starczak, V. R., DeCarlo, T. M. & Shamberger, K. E. F. (2015) Changes in coral reef communities across a natural gradient in seawater pH. *Science Advances*, **1**, DOI: 10.1126/sciadv.1500328

Birkeland, C. (2004) Ratcheting down the coral reefs. *BioScience*, **54**, 1021-1027.

Bonaldo, R. M., Welsh, J. Q. & Bellwood, D. R. (2012) Spatial and temporal variation in coral predation by parrotfishes on the GBR: evidence from an inshore reef. *Coral Reefs*, **31**, 263-272.

Bowden-Kerby, A. (2001) Low-tech coral reef restoration methods modelled after natural fragmentation processes. *Bulleting of Marine Science*, **69**, 915-931.

Bradbury, R. H. & Young, P. C. (1981) The effects of a major forcing function, wave energy, on a coral reef ecosystem. *Marine Ecology Progress Series*, **5**, 229-241.

Brander, L. M., Van Beukering, P. & Cesar, H. S. J. (2007) The recreational value of coral reefs: A meta-analysis. *Ecological Economics*, **63**, 209-218.

Bryan, T. L. & Metaxas, A. (2007) Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific Continental Margins of North America. *Marine Ecology Progress Series*, **330**, 113-126.

Caley, E. R. & Richards, J. F. C. (1956) *Theophrastus on stones*, 45-62. The Ohio State University, Columbus, U.S.A.

Carleton Ray, G. (1996) Coastal-marine discontinuities and synergisms: implications for biodiversity conservation. *Biodiversity & Conservation*, **5**, 1095-1108.

Chabanet, P., Ralambondrainy, H., Amaniue, M., Faure, G. & Galzin, R. (1997) Relationships between coral reef substrata and fish. *Coral Reefs*, **16**, 93-102.

Chefaoui, R. M. & Lobo, J. M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling,* **210,** 478–486.

Childs, C. (2004) Interpolating surfaces in ArcGIS spatial analyst. *ArcUser*, July-September, 32-35.

Cole, A. J., Lawton, R. J., Pratchett, M. S. & Wilson, S. K. (2011) Chronic coral consumption by butterflyfishes. *Coral Reefs*, **30**, 85-93.

Coles, S. L. (1976) A comparison of effects of elevated temperature versus temperature fluctuations on reef corals at Kahe Point, Oahu. *Pacific Science*, **29**, 15-18.

Coles, S. L. & Jokiel, P. L. (1977) Effects of temperature on photosynthesis and respiration in hermatypic corals. *Marine Biology*, **43,** 209-216.

Copper, P. (1994) Ancient reef ecosystem expansion and collapse. *Coral Reefs*, **13**, 3-11.

Couce, E., Ridgwell, A. & Hendy, E. J. (2012) Environmental controls on the global distribution of shallow-water coral reefs. *Journal of Biogeography*, **39**, 1508-1523.

Davis, G. E. (1977) Anchor damage to a coral reef on the coast of Florida. *Biological Conservation*, **11**, 29-34.

Davies, A. J., Wisshak, M., Orr, J. C. & Roberts, M. (2008) Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia). *Deep Sea Research Part I: Oceanographic Research Papers*, **55**, 1048-1062.

De Vooys, C. G. N. (1979) Primary production in aquatic environments. *The Global Carbon Cycle* (ed. by B. Bolin, E. T. Degens, S. Kempe & P. Ketner), Ch. 10, Scientific Committee on Problems of the Environment (SCOPE), Chichester, U.K.

Dubinsky, Z. & Stambler, N. (1996) Marine pollution and coral reefs. *Global Change Biology*, **2**, 511-526.

Elith, J. & Leathwick, J. R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics,* **40,** 677-697.

Fabricius, K. E. (2005) Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine Pollution Bulletin*, **50**, 125-146.

Ficetola, G. F., Thuiller, W. & Miaud, C. (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species—the American bullfrog. *Diversity and Distributions*, **13,** 476–485.

Fisher, R. L., Jantsch, M. J. & Comer, R. L. (1982) *General bathymetric chart of the oceans (GEBCO)*. Canadian Hydrographic Service, Ottawa, Canada.

Foster, K. A., Foster, G., Tourenq, C. & Shuriqi, M. K. (2011) Shifts in coral community structures following cyclone and red tide disturbances within the Gulf of Oman (United Arab Emirates). *Marine Biology*, **158,** 955-968.

Fricke, H. W. & Schuhmacher, H. (1983) The depth limits of Red Sea stony corals: an ecophysiological problem (a deep diving Survey by submersible). *Marine Ecology*, **4**, 163-194.

Goodman, J. A., Purkis, S. J. & Phinn, S. R. (2013) *Coral Reef Remote Sensing: A Guide for Mapping, Monitoring and Management*. Springer, Dordrecht, Germany.

Grigg, R. W. (2006) Depth limit for reef building corals in the Au'au Channel, S.E. Hawaii. *Coral Reefs*, **25**, 77-84.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8,** 993–1009.

Hughes, T. P., Baird, A. H., Dinsdale, E. A., Moltschaniwskyj, N. A., Pratchett, M. S., Tanner, J. E. & Willis, B. L. (2012) Assembly Rules of Reef Corals Are Flexible along a Steep Climatic Gradient. *Current Biology*, **22**, 736-741.

Jackson, J. B. C. (1992) Pleistocene perspectives on coral reef community structure. *American Zoologist*, **32**, 719-731.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An introduction to statistical learning*. Springer, New York, U.S.A.

Jokiel, P. L. & Coles, S. L. (1977) Effects of temperature on the mortality and growth of Hawaiian reef corals. *Marine Biology*, **43,** 201-208.

Jokiel, P. L. & York, R. H. (1982) Solar ultraviolet photobiology of the reef coral *Pocillopora damicornis* and symbiotic zooxanthellae. *Bulletin of Marine Science*, **32**, 301-315.

Kerry, J. T. & Bellwood, D. R. (2015) Do tabular corals constitute keystone structures for fishes on coral reefs? *Coral Reefs*, **34**, 41-50.

Kleypas, J. A., McManus, J. W. & Meñez, L. A. B. (1999) Environmental limits to coral reef development: where do we draw the line? *American Zoologist*, **39**, 146-159.

Kühl, M., Cohen, Y., Dalsgaard, T., Jørgensen, B. B. & Revsbech, N. P. (1995) Microenvironment and photosynthesis of zooxanthellae in scleractinian corals studied with microsensors for O_2, pH and light. *Marine Ecology Progress Series*, **117**, 159-172.

Leathwick, J. R., Elith, J. & Hastie, T. (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, **199,** 188–196.

León, Y. M. & Bjorndal, K. A. (2002) Selective feeding in the hawksbill turtle, an important predator in coral reef ecosystems. *Marine Ecology Progress Series*, **245**, 249-258.

Lough, J. M. & van Oppen, M. J. (2009) Introduction: coral bleaching – patterns, processes, causes and consequences. *Coral Bleaching* (ed. by J.M. Lough, M.J. van Oppen & M. Janice), pp. 1-5. Springer-Verlag, Berlin, Germany.

Lozano-Álvarez, E., Briones-Fourzán, P., Osorio-Arciniegas, A., Negrete-Soto, F. & Barradas-Ortiz, C. (2007) Coexistence of congeneric spiny lobsters on coral reefs: differential use of shelter resources and vulnerability to predators. *Coral Reefs*, **26**, 361-373.

Macintyre, I. G. & Pilkey, O. H. (1969) Tropical reef corals: tolerance of low temperatures on the North Carolina continental shelf. *Science*, **166**, 374-375.

Markager, S. & Vincent, W. F. (2000) Spectral light attenuation and the absorption of UV and blue light in natural waters. *Limnology and Oceanography*, **45**, 642-650.

Marubini, F. & Atkinson, M. J. (1999) Effects of lowered pH and elevated nitrate on coral calcification. *Marine Ecology Progress Series*, **188**, 117-121.

Mateo, R. G., Croat, T. B., Felicísimo, Á. M. & Muñoz, J. (2010) Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distribution*, **16,** 84–94.

McClanahan, T. R. (1994) Kenyan coral reef lagoon fish: effects of fishing, substrate complexity, and sea urchins. *Coral Reefs*, **13**, 231-241.

McCoy, J. & Johnston, K. (2001) *Using ArcGIS spatial analyst: GIS by ESRI*. Environmental Systems Research Institute, Redlands, U.S.A.

Meyer, C. G. & Holland, K. N. (2008) Spatial dynamics and substrate impacts of recreational snorkelers and SCUBA divers in Hawaiian Marine Protected Areas. *Journal of Coastal Conservation*, **12**, 209-216.

Miller, C. B. & Wheeler, P. A. (2012) *Biological Oceanography*, 2nd ed. Oregon State University, Corvallis, U.S.A.

Milliman, J. D. (1974) *Marine Carbonates*. Springer-Verlag, Berlin, Germany.

Moberg, F. & Folke, C. (1999) Ecological goods and services of coral reef ecosystems. *Ecological Economics*, **29**, 215-233.

Mumby, P. J., Green, E. P., Edwards, A. J. & Clark, C. D. (1997) Coral reef habitat mapping: how much detail can remote sensing provide? *Coral Reefs*, **130**, 193-202.

Mumby, P. J., Skirving, W., Strong, A. E., Hardy, J. T., LeDrew, E. F., Hochberg, E. J., Stumpf, R. P. & David, L. T. (2004) Remote sensing of coral reefs and their physical environment. *Marine Pollution Bulletin*, **48**, 219-228.

Muthiga, N. A. & Szmant, A. M. (1987) The effects of salinity stress on the rates of aerobic respiration and photosynthesis in the hermatypic coral *Siderastrea siderea*. *Biological Bulletin*, **173**, 539-551.

NASA. (2007) Millennium coral reefs landsat archives. *USF Millennium Global Coral Reef Mapping Project*, St. Petersburg, U.S.A., http://oceancolor.gsfc.nasa.gov/cgi/landsat.pl

Newell, N. D. (1971) An outline history of tropical organic reefs. *American Museum Novitates*, **2465**, 1-17.

NOAA. (2013) World Ocean Database. *National Centers for Environmental Information*, http://www.nodc.noaa.gov/OC5/WOD/pr_wod.html

Osborne, K., Dolman, A. M., Burgess, S. C. & Johns, K. A. (2011) Disturbance and the dynamics of coral cover on the Great Barrier Reef (1995–2009). *PLOS ONE*, **9**(6): e99742, DOI: 10.1371/journal.pone.0017516

Palacios, M. M., Muñoz, C. G. & Zapata, F. A. (2014) Fish corallivory on a pocilloporid reef and experimental coral responses to predation. *Coral Reefs*, **33**, 625-636.

Parrish, F. A., Craig, M. P., Ragen, T. J., Marshall, G. J. & Buhleier, B. M. (2000) Identifying diurnal foraging habitat of endangered Hawaiian monk seals using a seal-mounted video camera. *Marine Mammal Science*, **16**, 392-412.

Polónia, A. R. M., Cleary, D. F. R., de Voogd, N. J., Renema, W., Hoeksema, B. W., Martins, A. & Gomes, N. C. M. (2015) Habitat and water quality variables as predictors of community composition in an Indonesian coral reef: a multi-taxon study in the Spermonde Archipelago. *Science of the Total Environment*, **537**, 139-151.

ReefBase (2016) A global information system for coral reefs. *WorldFish*, http://www.reefbase.org

Spalding, M. D. & Grenfell, A. M. (1997) New estimates of global and regional coral reef areas. *Coral Reefs*, **16**, 225-230.

Spalding, M. D., Ravilious, C. & Green, E. P. (2001) *Atlas of Coral Reefs*. The University of California Press, Berkeley, U.S.A.

Stoddart, D. R. (1971) Coral reefs and islands and catastrophic storms. *Applied Coastal Geomorphology* (ed. by J. A. Steers), pp. 155-197. Palgrave Macmillan, London, U.K.

R Development Core Team (2016) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/

Smith, S. V. (1978) Coral-reef area and the contributions of reefs to processes and resources of the world's oceans. *Nature*, **273**, 225-226.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, **21**, 272-281.

UNEP-WCMC (2010) *Global distribution of warm-water coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project*. The United Nations Environment Programme, http://data.unep-wcmc.org/datasets/1 (2010)

Walbran, P. D., Henderson, R. A., Jull, A. J. T. & Head, M. J. (1989) Evidence from sediments of long-term *Acanthaster planci* predation on corals of the Great Barrier Reef. *American Association for the Advancement of Science*, **245**, 847-850.

Wolf, A. T. & Nugues, M. M. (2013) Predation on coral settlers by the corallivorous fireworm *Hermodice carunculata*. *Coral Reefs*, **32**, 227-231.

Wood, S. N. (2001) mgcv: GAMs and generalized ridge regression for R. *R News*, **1,** 20–25.

Zuur, A., Ieno, E. N. & Smith, G. M. (2007) *Analysing ecological data*. Springer Science & Business Media, New York, U.S.A.

Figures

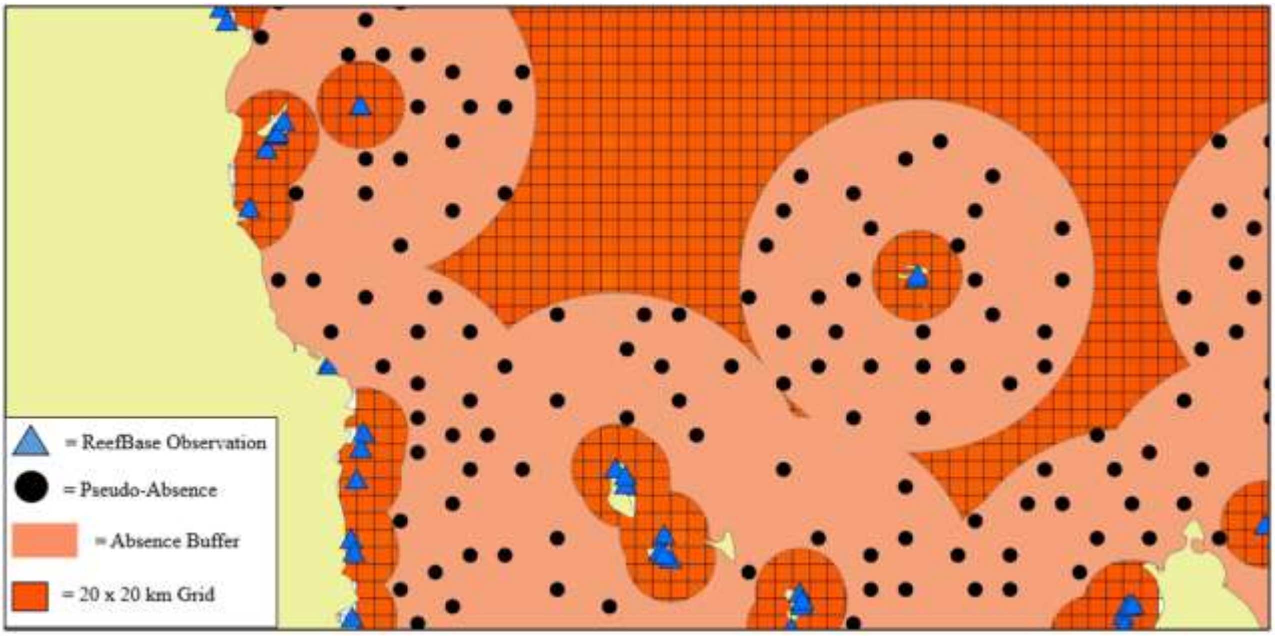

Figure 1: A visualization of how tropical coral reef presences and absences were determined. Pseudo-absences were generated on the basis of 40 – 200 km radial buffers, and were randomly distributed within them.

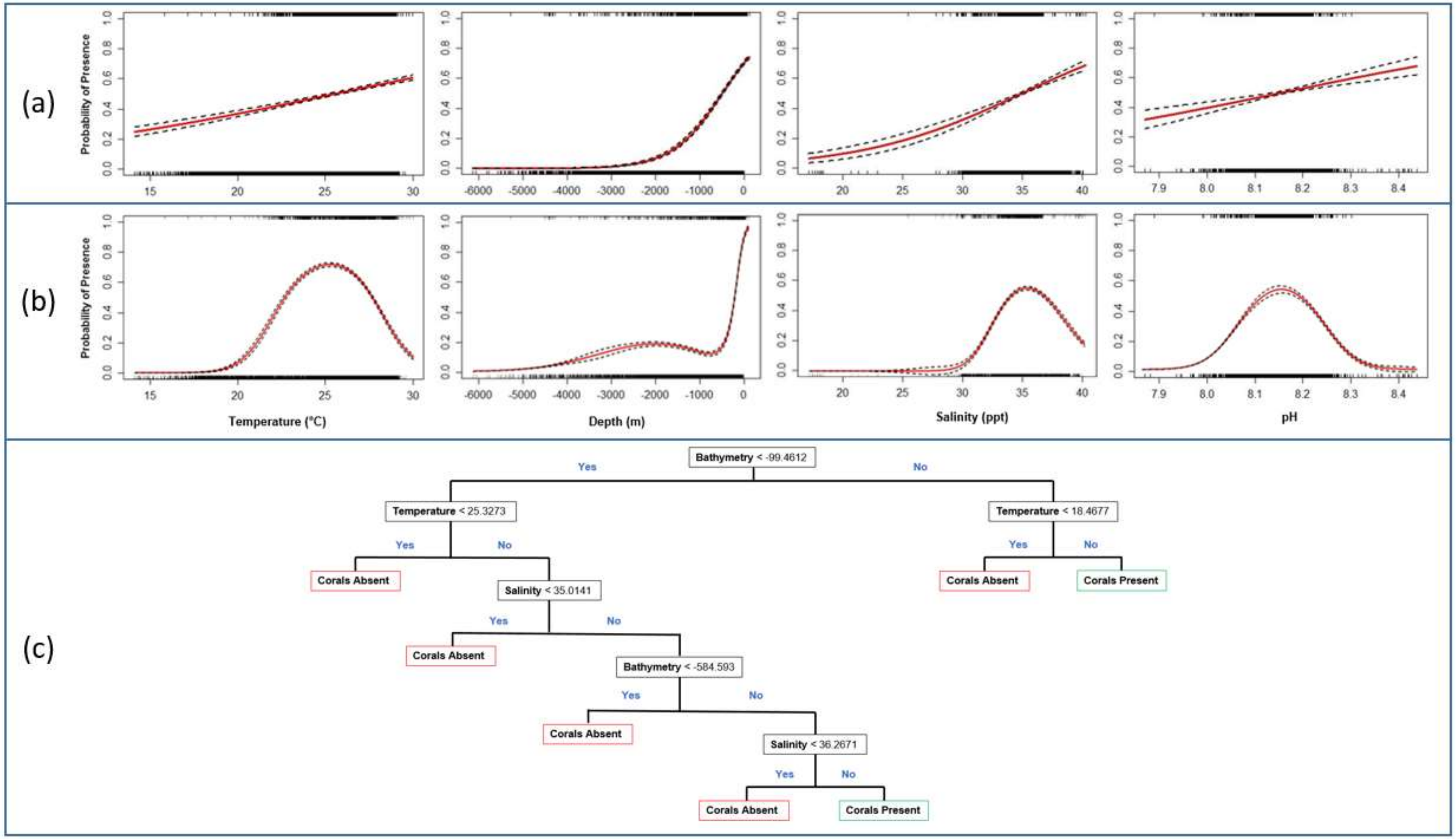

Figure 2: Marginal effects of the four environmental parameters, on the probability of observing a tropical coral reef. Panel A shows the effects, as they were observed in the generalized linear model (GLM), while Panel B shows the effects within the generalized additive model (GAM). Panel C shows the effects of the environmental parameters on the structure of the classification and regression tree (CRT). Dashed lines represent 95% confidence intervals. Tick marks along the secondary x-axis represent presence values, while tick marks along the x-axis represent pseudo-absences.

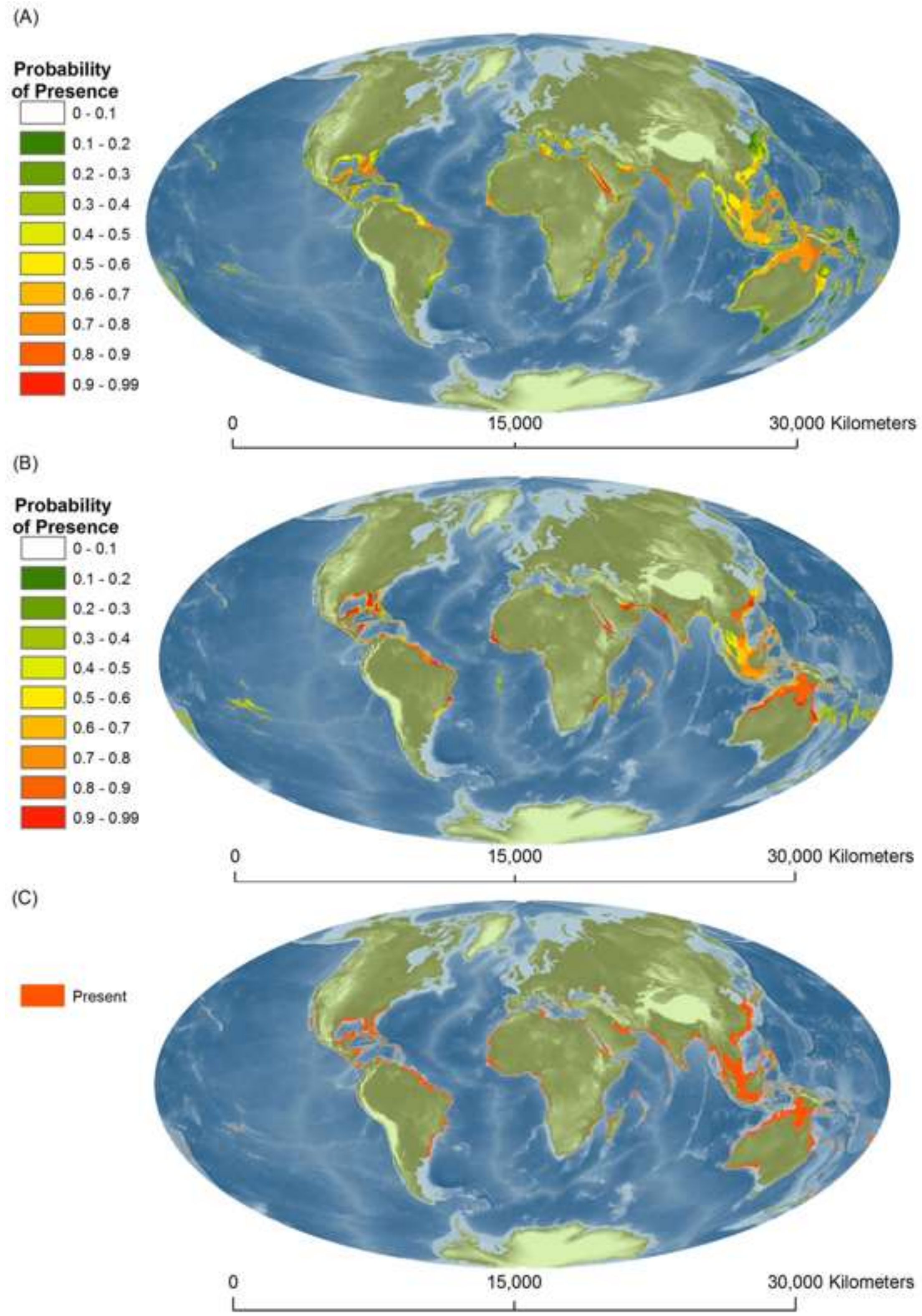

Figure 3: A visual comparison of the three model outputs. Panel A displays the predictions of the generalized linear model (GLM). Panel B shows the predictions of the generalized additive model (GAM), while Panel C shows the output of the classification and regression tree (CRT). CRT model outputs were binary; thus, coloration indicates presence, and no coloration indicates an absence of suitable habitat for tropical coral reefs. Model outputs are presented as Aitoff projections.

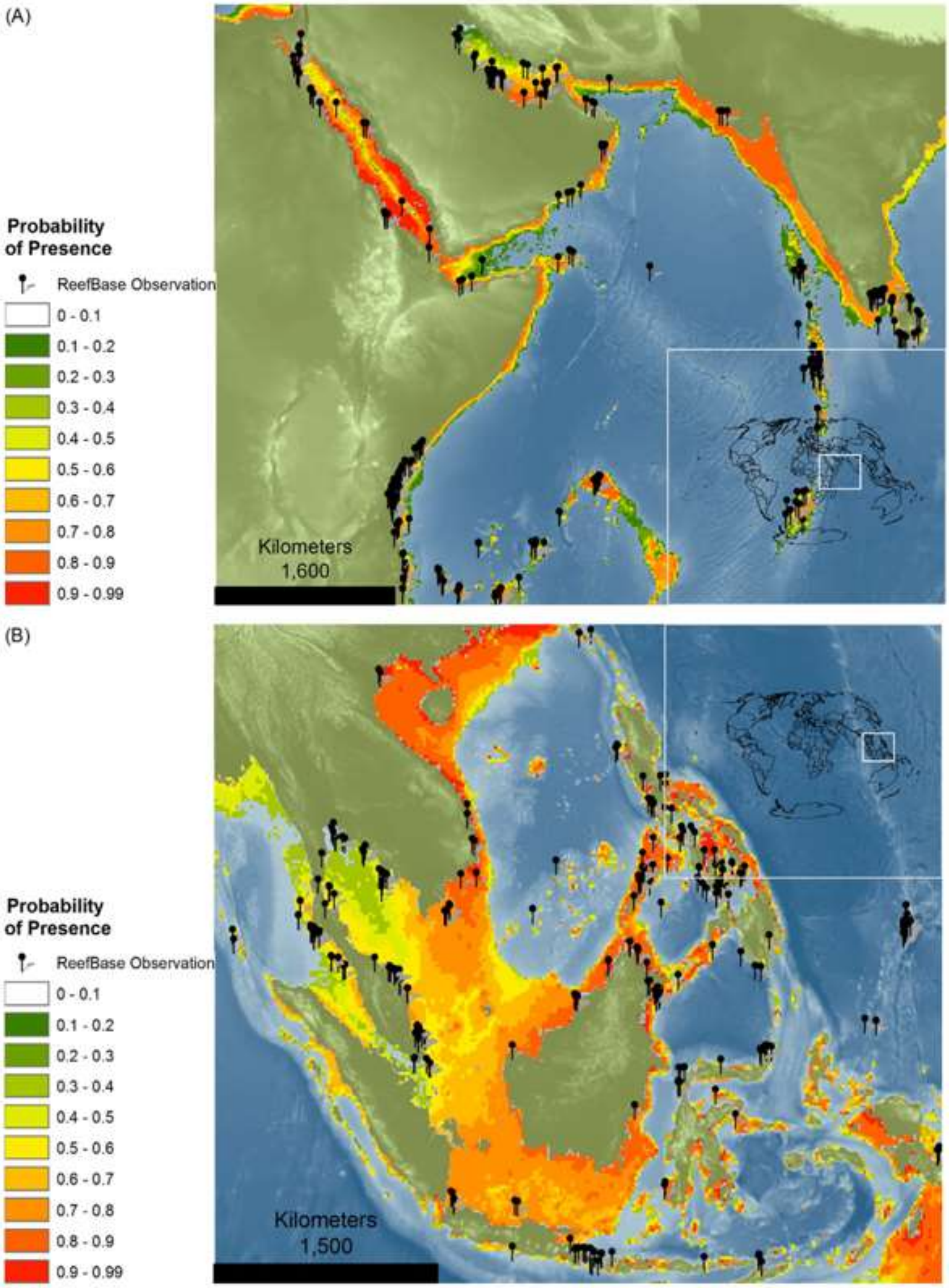

Figure 4: Locations of previous observations of coral reef presence, as they relate to the predicted areas of suitable habitat for tropical coral reefs. Panels A and B display the outputs of the generalized linear model in the Indian and Pacific Ocean, respectively. Model outputs are presented as Aitoff projections.

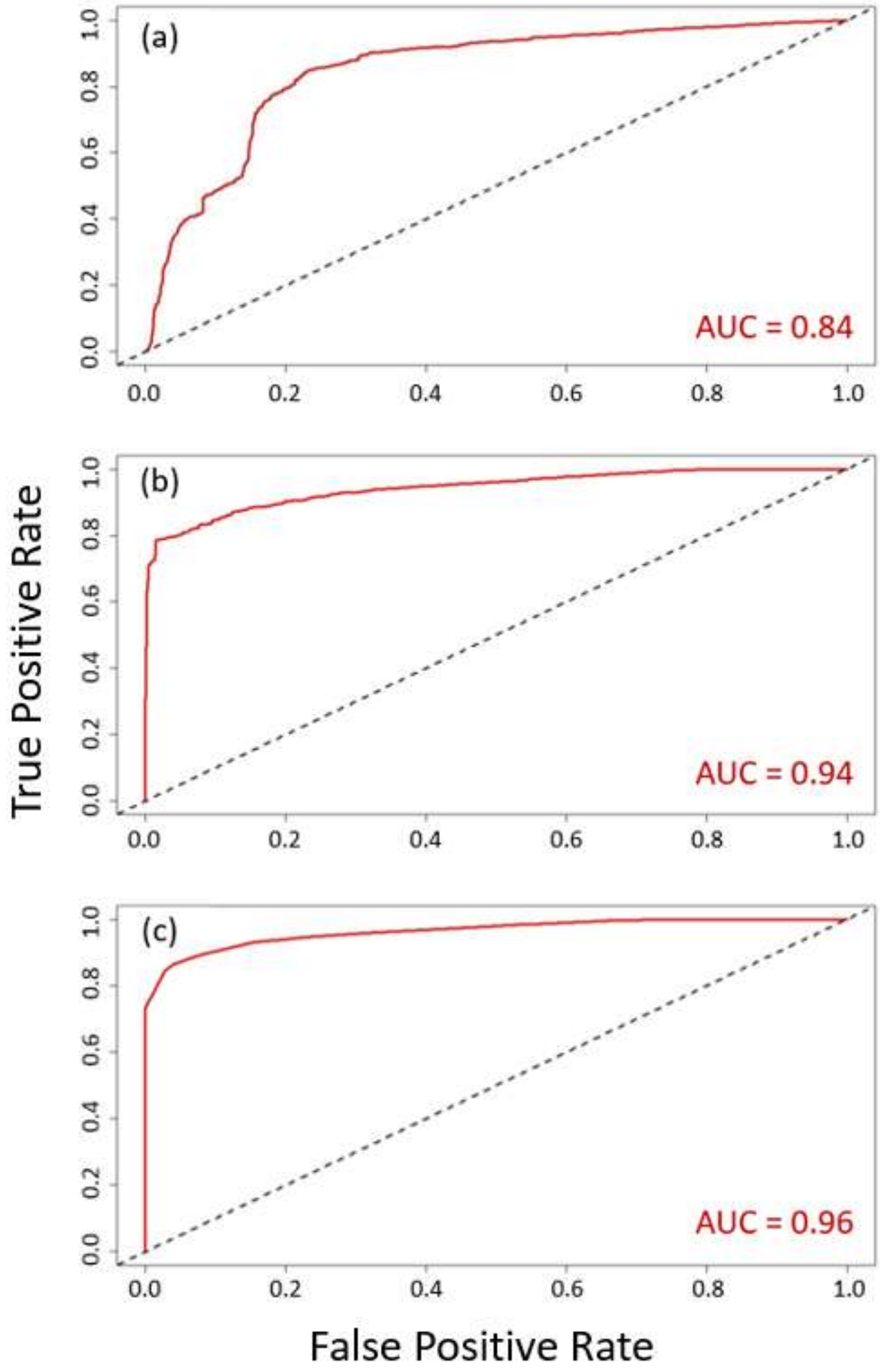

Figure 5: Receiver Operating Characteristic (ROC) curves of the three statistical models used to predict suitable habitat for hermatypic corals; which was used as a proxy for coral reef habitat. Panel A shows the ROC curve and associated AUC value of the generalized linear model (GLM). Panels B and C show the generalized additive model (GAM) and the classification and regression tree (CRT), respectively.