

RESEARCH ARTICLE

Models of Marine Fish Biodiversity: Assessing Predictors from Three Habitat Classification Schemes

Katherine L. Yates^{1,2*}, Camille Mellin^{1,3}, M. Julian Caley¹, Ben T. Radford⁴, Jessica J. Meeuwig⁵

1 Australian Institute of Marine Science, PMB 3, Townsville, Queensland, Australia, **2** School of Environment and Life Sciences, University of Salford, Manchester, United Kingdom, **3** Global Change Ecology Lab, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia, **4** Australian Institute of Marine Science, UWA Oceans Institute, Crawley, Western Australia, Australia, **5** Centre for Marine Futures, Oceans Institute and School of Animal Biology, University of Western Australia, Crawley, Western Australia, Australia

* k.l.yates@salford.ac.uk



OPEN ACCESS

Citation: Yates KL, Mellin C, Caley MJ, Radford BT, Meeuwig JJ (2016) Models of Marine Fish Biodiversity: Assessing Predictors from Three Habitat Classification Schemes. PLoS ONE 11(6): e0155634. doi:10.1371/journal.pone.0155634

Editor: Jan Geert Hiddink, Bangor University, UNITED KINGDOM

Received: May 13, 2015

Accepted: May 2, 2016

Published: June 22, 2016

Copyright: © 2016 Yates et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data are available through the Australian Ocean Data Network portal (ODN) and are publically available, <http://portal.aodn.org.au/aodn/>.

Funding: No specific funding was obtained for this piece of research. However, the original data collection was funded by the Natural Heritage Trust II (as part of a different project) and CM is funded by an Australian Research Council grant (DE140100701).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Prioritising biodiversity conservation requires knowledge of where biodiversity occurs. Such knowledge, however, is often lacking. New technologies for collecting biological and physical data coupled with advances in modelling techniques could help address these gaps and facilitate improved management outcomes. Here we examined the utility of environmental data, obtained using different methods, for developing models of both uni- and multivariate biodiversity metrics. We tested which biodiversity metrics could be predicted best and evaluated the performance of predictor variables generated from three types of habitat data: acoustic multibeam sonar imagery, predicted habitat classification, and direct observer habitat classification. We used boosted regression trees (BRT) to model metrics of fish species richness, abundance and biomass, and multivariate regression trees (MRT) to model biomass and abundance of fish functional groups. We compared model performance using different sets of predictors and estimated the relative influence of individual predictors. Models of total species richness and total abundance performed best; those developed for endemic species performed worst. Abundance models performed substantially better than corresponding biomass models. In general, BRT and MRTs developed using predicted habitat classifications performed less well than those using multibeam data. The most influential individual predictor was the abiotic categorical variable from direct observer habitat classification and models that incorporated predictors from direct observer habitat classification consistently outperformed those that did not. Our results show that while remotely sensed data can offer considerable utility for predictive modelling, the addition of direct observer habitat classification data can substantially improve model performance. Thus it appears that there are aspects of marine habitats that are important for modelling metrics of fish biodiversity that are not fully captured by remotely sensed data. As such, the use of remotely sensed data to model biodiversity represents a compromise between model performance and data availability.

Introduction

Globally, marine biodiversity provides myriad and valuable ecosystem goods and services [1–3]. These goods and services, however, are threatened by an extensive range of natural and anthropogenic stressors, many of which are broadly distributed [4,5]. Consequently, many of these goods and services are in decline [2]. The conservation and sustainable management of marine biodiversity is hampered by insufficient information [5] and taxonomic bias in the information currently available [6]. Effective conservation and management is further impeded by the difficulty and cost of acquiring additional data from marine habitats due to their extent and inaccessibility, as well as the complexity of these systems. The lack of effective biological surrogates [7,8], and the often insufficient opportunities to supplement current information, means that other approaches must be used to overcome these information and knowledge gaps.

Recent advances in collecting and analysing marine data are supporting new analytical techniques that can help fill these gaps, and in doing so, contribute to more effective conservation and management of living marine resources [9,10]. Tools such as baited cameras and towed video enable direct observation of marine species and their habitats, in more affordable and efficient ways, and in places divers cannot access [11–13]. Multibeam sonar is also now commonly used to generate high-resolution bathymetry of marine habitats (e.g. [10]). In turn, such bathymetry supports habitat classification at very fine spatial resolution over large areas of sea-floor [10,14,15].

More commonly now, predictive biodiversity models developed with remotely sensed data are being developed and applied to marine ecosystems. However, whether predicted habitat classification provides better predictors for modelling metrics of biodiversity (e.g. species richness) than predictors taken directly from remotely sensed data is not yet known. Similarly, predictors from remotely sensed data have not previously been compared to predictors obtained from direct observer habitat classification. Assessment of the strength of predictors produced by these three methods could improve our understanding of how to build better models by informing the selection of sampling methods, and the suitability of such models according to their intended use and the relative resource requirements of information acquisition. Furthermore, the use of such models to predict assemblage-level metrics has, so far, been largely limited to predicting total species richness, total abundance or total biomass (e.g. [16]). Utility of these models would be enhanced if they could predict a wider suite of metrics that are informative with respect to conservation and management objectives. Such metrics might include abundance and biomass of specific groups such as vulnerable species or endemic species, or functional group membership more broadly.

Here we investigate which of a range of habitat classification methods provides the most powerful predictors when modelling metrics of fish biodiversity and explore which of a series of metrics can be most effectively modelled. We use data from the Marine Futures Project (2005–2008), funded through Australia's Natural Heritage Trust. Using these data, we test the predictive power of environmental parameters obtained in three different ways: acoustic (multibeam) sonar imagery, predicted habitat classification, and direct observer habitat classification. We then investigate the relative influence of individual predictors and test the ability of models to predict a series of community-level univariate and multivariate metrics: total species richness, total abundance, total biomass, abundance and biomass of vulnerable species ('vulnerability', after [17]), abundance and biomass of fisheries targeted species, abundance and biomass of endemic species, and functional group composition by abundance and biomass.

Methods

Study Site

Rottnest Island (Fig 1, Geographic Datum of Australia 94 Zone 50 South, GDA 94 Z50) is biologically diverse, and includes a wide range of habitats from tropical coral reefs to rocky temperate reefs, seagrass beds and sandy barrens. This diversity reflects the strong influence of the Leeuwin Current, a south-flowing boundary current that brings warm tropical water and species into this otherwise sub-tropical region. As such, the region supports an unusual mix of tropical and temperate fishes [18]. Rottnest Island and its environs are also an important recreational area in close proximity to Western Australia's state capital, Perth, and as such, marine based tourism and recreational activities including sailing, diving and fishing are highly valued [19]. There is limited commercial fishing in the area and no oil and gas exploration activity. Despite the area's high conservation and socioeconomic value, there is little information on the distribution of its marine biodiversity, and therefore, the implication of these distributions for management.

Rottnest Island was surveyed in 2007 as part of the Marine Futures Program (www.marinefutures.fnas.uwa.edu.au). Surveys included multibeam mapping, towed video and 349 Baited Remote Underwater Video (BRUV) deployments. From these surveys the species data (fish) and habitat data (three types) used in this study were obtained. These data and the methods used to obtain them are described below.

Habitat data

An area around Rottnest (~250 km²) was subject to a full coverage multibeam survey conducted by Fugro (PTY). The multibeam survey was conducted using a Reson 8101 Multibeam. The multibeam sensor point cloud varied with water depth but on average collected one point or ping every ~50-100cm, which was averaged over a 2.5m cell. Development of secondary datasets from the hydroacoustic data provided textural information about the seafloor. Data were processed to construct full coverage bathymetry maps with a 2.5m cell size. Combined, these data formed our first group of (27) habitat variables, hereafter referred to as multibeam habitats.

Approximately 100 km of towed video imagery was also collected. Processing of this imagery consisted of classifying the observed imagery with respect to habitat type and geo-referencing position such that boundaries between habitat types, and thus patches of habitat types, were identified ([20]). Habitat types were classified by trained and cross-validated analysts into 34 broad classes using a standard classification scheme for Australia [21]. The multibeam and the habitat classification data from the towed video were used to generate predicted habitat classifications across the study area, again at a cell size of 2.5m. Regression trees [22,23] were used to predict the probability of the presence of a given habitat type in a given cell (See [S1 File](#), supplementary methods, for full details). These data formed our second group of (34) habitat variables, hereafter referred to as predicted habitats.

Lastly, during the analysis of the BRUVS images, habitats observed were classified by visual observation in to one of five abiotic habitat classes (high, medium and low profile reef, sand inundated reef, or sand) and one of six biotic habitat classes (macroalgae, seagrass, coral, sessile invertebrates, bare). These data formed our third group of (2) categorical habitat variables (abiotic and biotic categories), hereafter referred to as direct observer habitats.

Species Data

The predicted habitat map generated from the multibeam and towed video formed the basis of the sampling design for the fish survey using stereo-BRUVS. A total of 349 BRUV deployments

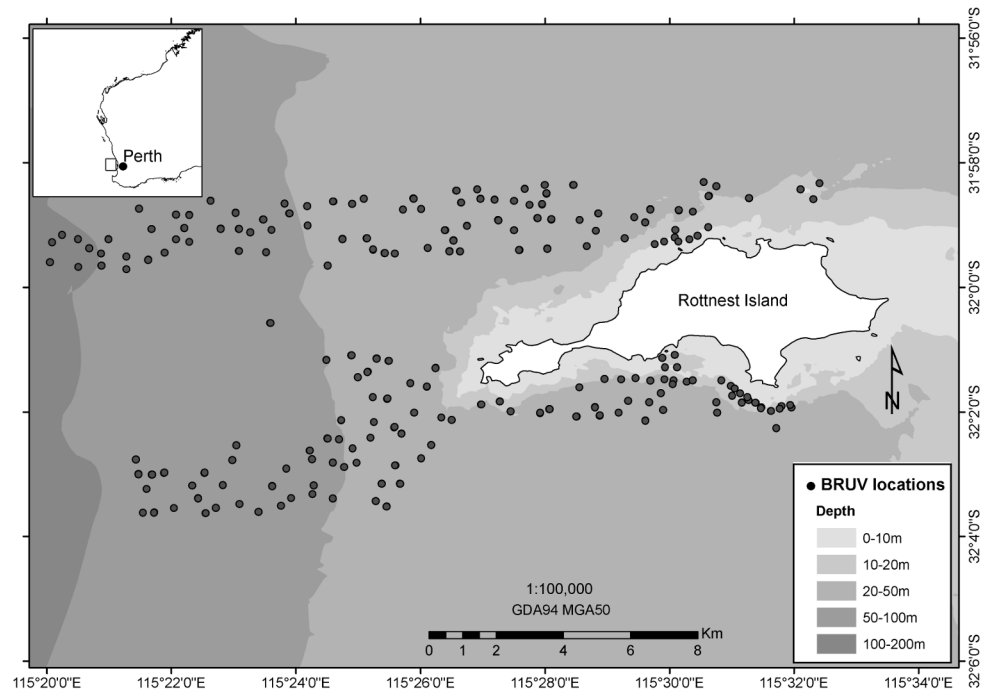


Fig 1. Map of Rottnest study site, Western Australia. Geographic Datum of Australia 94 Zone 50 South (GDA 94 Z50).

doi:10.1371/journal.pone.0155634.g001

(hereafter referred to as samples) were made at the Rottnest site. All samples with visibility ≤ 3 m were excluded, leaving a total of 280 samples used in this study. From each of the remaining samples, fish species were identified and their relative abundances and individual fork lengths were estimated using Event-Measure [24]. See S1 File, supplementary methods, for full details of BRUVS sampling design and image analysis.

A series of biodiversity metrics were calculated for each sample (Table 1). These metrics were derived from species identifications, relative species abundances and individual fork lengths. Biomass was estimated by calculating individual fish weight from fork length using relationships available on FishBase [25]. Fishbase was also used: to assign each species to a trophic level and score their vulnerability to exploitation [17], to class each species as either

Table 1. Estimates of biodiversity metrics for marine fishes around Rottnest Island, Western Australia. Vulnerability, target and endemic metrics are reported as a percentage of the total abundance or total biomass. Results are from 280 samples, each sample being an individual BRUVs deployment.

Metric	Mean	Std.Dev	min	max
Total species richness	10.8	5.8	1	31
Total abundance	55.5	75.2	1	1041
Total biomass (g)	33252.1	51675.2	67.9	735990.2
Vulnerability (abundance)	4.3	8.9	16.9	87
Vulnerability (biomass)	61.8	11.4	24.2	87
Target (abundance)	54.1	26.8	0	100
Target (biomass)	77.6	23.1	0	100
Endemic (abundance)	83.1	21.7	0	100
Endemic (biomass)	48	31.9	0	100

doi:10.1371/journal.pone.0155634.t001

endemic to Australia or not, to assign each species to a functional group, and to ascertain whether or not each species was targeted by fisheries [25].

Vulnerability estimates were based on species' life history traits and scored on a scale of 1 (lowest) to 100 (highest) [17]. Classification to functional group was based on information on prey items, trophic level and maximum body size, as reported in Fishbase (see [S1 File](#), supplementary methods).

The large number of benthic omnivores and zoobenthivores and the large variations in their body sizes led us to separate them into small and large size classes within both groups of species. Small and large omnibenthivores were separated at a mean observed size of 31.5 cm while small and large zoobenthivores were separated at a mean observed size of 35.0 cm. The classification of individuals into small and large categories for the omnibenthivores and zoobenthivores was based on the distribution of observed sizes, with the former trophic category tending towards smaller individuals.

Modelling

Variable selection. Our set of potential predictor variables was large and included estimates of 27 multibeam derived variables and 34 predicted habitat variables, plus two direct observer habitat variables (abiotic & biotic categories) and depth (see [S1 Table](#) for a detailed description). Whilst regression trees largely ignore non-informative predictors when fitting trees [26,27], predictor selection is still useful because redundant predictors can degrade a model's performance by increasing variance (Elith *et al* 2008). BRT are also only robust to moderate levels of multicollinearity [28]. Therefore, within each of the multibeam and predicted habitat variable sets, we selected predictors for use in our models based on three steps. We identified predictors that were highly correlated (Pearson's $r > 0.8$). We developed an initial set of exploratory BRT models including all candidate predictors and examined their relative influence, identifying candidate predictors for which the relative influence was consistently $\leq 1\%$. Finally, we ran a step-wise simplification of those exploratory models, using methods analogous to backward selection in regression (after Elith *et al* 2008), to see which variables were repeatedly dropped from the models. Based on our initial investigation of these candidate predictors, we reduced our predictors to 12 multibeam variables, 11 predicted habitat variables, the two direct observer habitat variables (abiotic and biotic categories) and depth ([S4 Table](#)).

Boosted regression trees. We used boosted regression trees (BRT) (Elith *et al.* 2006, 2008) to model relationships between environmental characteristics and nine univariate, community level metrics ([Table 1](#)). Models were developed in R 3.1.1 [29], using the `gbm` package.

Distributions of community metrics were skewed. Therefore, we either $\log(x+1)$ transformed them if variables were null or positive (species richness, total abundance and total biomass) or logit transformed them if bounded between 0 and 1 (vulnerability, target and endemic metrics) to approximate a normal distribution [26,30]. For each community metric, models were developed for five sets of predictors: 1) multibeam plus direct observer habitats, 2) predicted habitats plus direct observer habitats, 3) multibeam, 4) predicted habitats, 5) direct observer habitats. All models also included depth.

Models were developed using a randomly selected 50% of the data (0.5 bag fraction). Following Elith *et al.* (2008), the 50% level was chosen based on model performance after a range of values between 30 and 70% were tested on an initial set of models, and their predictive powers compared. Cross-validation (CV), using bootstrapped data, was used to evaluate the predictive capacity of the models. For each combination of community level metric and set of predictors, a range of values for the two main model parameters, tree complexity and learning

rate, were explored and the best model selected based on the greatest CV deviance explained and lowest mean prediction error, following [26] (see [S2 Table](#) for the parameters of the best models). Model residuals were tested (Moran's I) to rule out spatial autocorrelation. In some cases, intra-model variability was of similar magnitude to inter-model variability, so each model was run five times with the same parameters and the average CV deviance and prediction error were calculated across model runs ([S3 Table](#) shows the ranges in CV deviance and prediction error for the best models). We also examined the relative influence (%) of each predictor on the best models (Figs [2](#) and [3](#)).

Multivariate regression trees. We used multivariate regression trees (MRT) (De'Ath 2002) to model the relationship between predictors and the relative abundance (or biomass) of the different trophic and vulnerability-based functional groups (R package mvpart). We defined groups by combining diet and vulnerability, which resulted in 29 functional groups ranging from 'high vulnerability piscivores' to 'low vulnerability zooplanktivores'. We built an initial model using predictors from the multibeam data, which we then compared to a second model built using predicted habitat classes derived from the multibeam—both models also included the direct observer habitats (biotic and abiotic categories) and depth.

MRT clusters sites by repeated splitting of the data, with each split determined by habitat characteristics [[31](#)] and corresponding to a distinct species assemblage. Tree fit is defined by the relative error (RE; total impurity of the final tree divided by the impurity of the original data). RE is an over-optimistic estimate of tree accuracy, which is better estimated from the cross-validated relative error (CVRE). We determined the best tree size (i.e. number of leaves or clusters formed by the tree) as that which minimized CVRE, which varies from zero for a perfect predictor to nearly one for a poor predictor (De'Ath 2002). We then examined the splits and quantified the variance that each of them explained, based on the entire dataset and for each individual functional group.

We identified functional groups and species that characterized each resulting cluster using the Dufrene-Legendre index, which is based on the relative abundance and frequency of each species (or functional group) within a given cluster [[32](#)]. The index varies between 0, no occurrences of a species within a cluster, to 100, if a species occurs at all sites within the cluster and in no other cluster. The index is associated with the probability of resulting from a random pattern, based on 250 reallocations of sites among clusters [[32](#)].

Ethical approval

Quantitative measurements of the fish assemblage were made using non-destructive sampling methods (BRUVS) and there was no sacrifice or incident mortality associated with this sampling method. The use of BRUVS for collecting the data used here required ethics approval, which was from the University of Western Australia and approval from the Western Australian Department of Fisheries. No other approvals were necessary, nor did sampling occur within protected areas or on private property. The field work did not impact any listed protected species.

Results

Boosted regression trees

BRT explained up to 63.9% CV deviance in response variables, depending on the set of predictors used and the community metric being modelled ([Table 2](#)). The best performing models were for total species richness (63.9% CV deviance explained) and total abundance (43.4% CV deviance explained). Biomass proved more difficult to model, with CV deviance explained consistently lower for biomass than for the corresponding abundance ([Table 2](#)). Abundance and

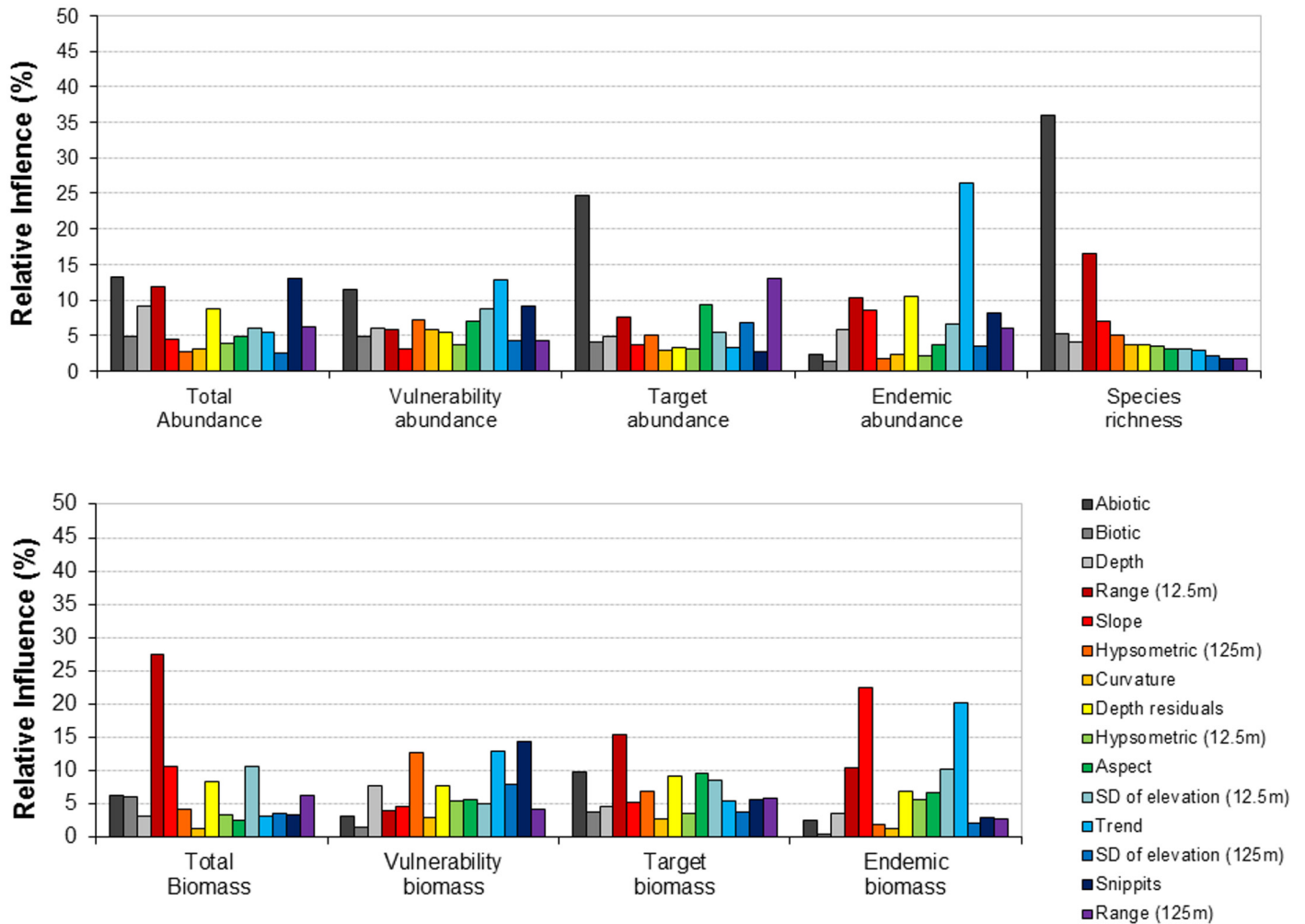


Fig 2. The relative influence (%) of individual predictors estimated using boosted regression tree models for nine fish diversity metrics. Models were developed with predictors from Multibeam data plus direct observer habitats (abiotic and biotic) and depth. Multibeam data and depth were numeric values, the abiotic and biotic variables were categorical (five abiotic and six biotic classes). Vulnerability, target and endemic metrics were developed for both the percentage of the total abundance and the percentage of the total biomass.

doi:10.1371/journal.pone.0155634.g002

biomass of endemic species proved the most difficult to model, with a CV deviance explained of only 10.0% and 8.4%, respectively.

Models developed using predictors from predicted habitats, plus direct observer habitats and depth generally had the greatest CV deviance explained, outperforming models developed with other sets of predictors for five out of nine metrics (Table 2). Models that excluded the direct observer habitats consistently performed worse than corresponding models including them, highlighting the importance of these two variables. Indeed, models including depth and the direct observer habitats alone were able to explain >80% of the maximum CV deviance for five of the nine community metrics, and >50% for eight of the nine metrics (across all metrics, range = 35–95%, mean = 70%).

Models without direct observer habitats also performed worse in terms of mean prediction error than corresponding models that included them (Table 2). In almost all cases, models for biomass had greater mean prediction error than the corresponding models for abundance, again indicating that biomass was less predictable than abundance. Models developed using

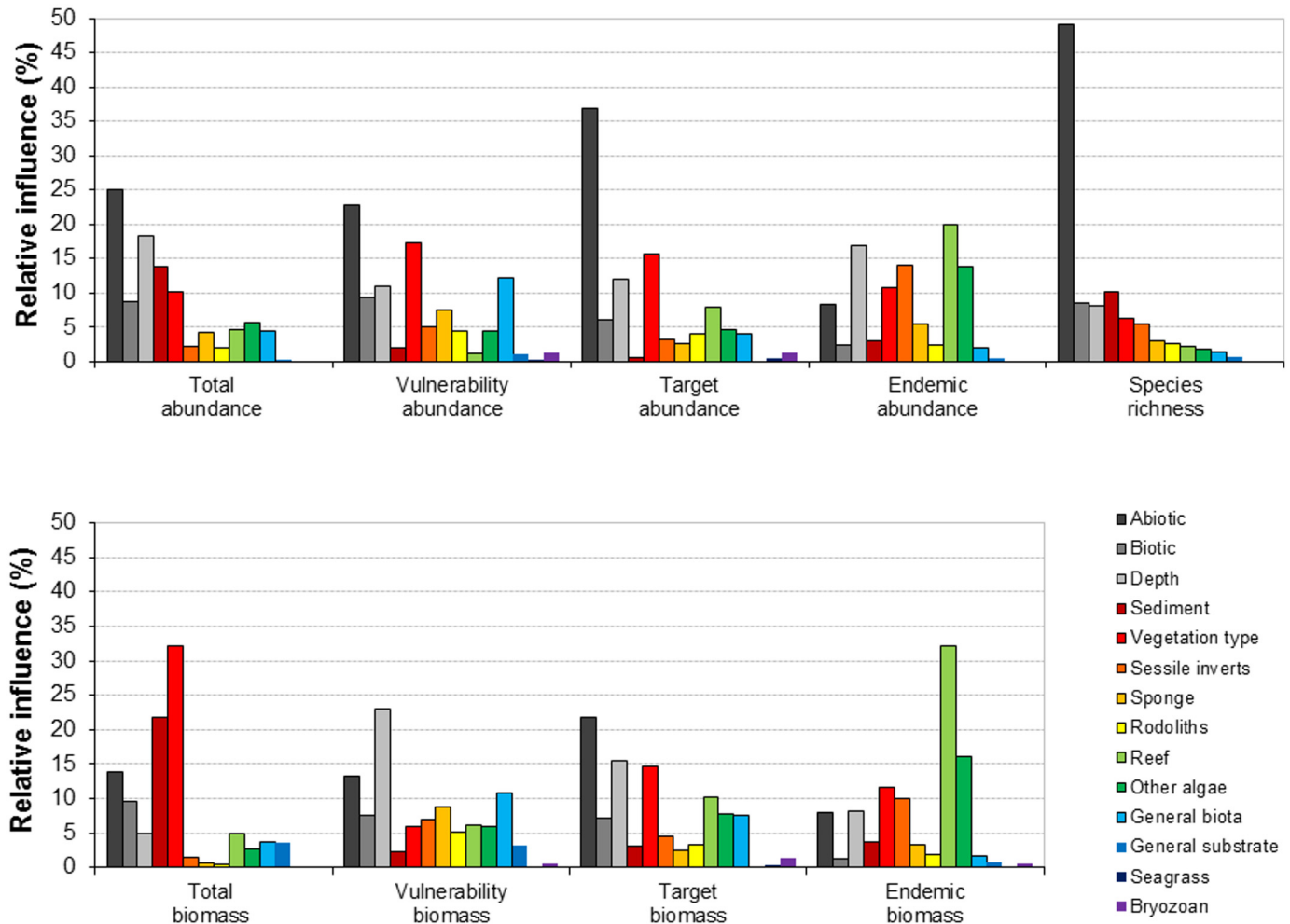


Fig 3. The relative influence (%) of individual predictors as estimated using boosted regression tree models for nine fish diversity metrics. Models were developed with predictors from predicted habitat classes, plus direct observer habitats (abiotic and biotic) and depth. Predicted habitat variables are numeric and describe the proportion of that habitat within the sample area, the abiotic and biotic variables were categorical (five abiotic and six biotic classes). Vulnerability, target and endemic metrics were developed for both the percentage of total abundance and the percentage of the total biomass.

doi:10.1371/journal.pone.0155634.g003

multibeam data had consistently lower prediction error than those developed with predicted habitats, however, the difference was generally very small (i.e. ≤ 0.01 for five of the nine metrics; Table 2).

When using either multibeam or predicted habitats, plus direct observer habitats and depth, the abiotic variable from the direct observer habitats generally had the greatest relative influence on BRTs, with 22.1% average influence on models developed using habitats and 12.2% on models developed using multibeam (Figs 2 and 3). Relative influence was much more evenly distributed between individual predictors for models developed using multibeam habitats than those developed from predicted habitats (Figs 2 and 3). The relative influence of individual predictors varied greatly across community metrics; the abiotic category for example had a much greater influence on abundance metrics ($28.5\% \pm 13.4$ when using habitats and $17.6\% \pm 11.5$ when using multibeam; mean \pm SD) than it did on biomass metrics ($14.2\% \pm 13.6$ when using habitats and $5.3\% \pm 11.3$ when using multibeam).

Table 2. Results of boosted regression tree models for nine biodiversity metrics, developed with five sets of predictors. Predictors were developed from three sets of habitat data: multibeam habitats (multibeam), predicted habitats (pred. habitats) and direct observer habitats (biotic and abiotic categories). All models also contained depth as a predictor. Vulnerability, target and endemic metrics were developed for both the percentage of the total abundance and the percentage of the total biomass. The cross validation (CV) deviance explained (%) and mean prediction error (%) are averaged values from five model runs. The greatest CV deviance explained and the lowest prediction error for each metric are highlighted in bold. Data were collected around Rottnest Island, Western Australia, in 2007.

	Metrics	Predictors				
		Multibeam, depth, biotic & abiotic	Pred. habitats, depth, biotic & abiotic	Multibeam & depth	Pred. habitats & depth	Biotic, abiotic & depth
CV deviance explained (%)	Total species richness	63.9	63.3	55.8	49.9	58.5
	Total abundance	43.4	42.1	38.9	34.6	39.6
	Total biomass	16.6	17.6	14.8	16.1	14.1
	Vulnerability (abundance)	25.9	27	22.1	23.5	23.4
	Vulnerability (biomass)	0.5	1.9	0.4	1.2	1
	Target (abundance)	26.9	26.9	19	11.8	24.3
	Target (biomass)	12.1	9	8.5	3	7
	Endemic (abundance)	8.5	10	8.2	9.3	5.5
	Endemic (biomass)	6.3	8.4	6.2	8.3	2.7
Mean prediction error (%)	Total species richness	10.4	11.9	11.5	14.6	13.4
	Total abundance	13.4	15.4	17	19.8	18.2
	Total biomass	8.1	8.5	8.3	8.6	8.7
	Vulnerability (abundance)	10.3	11.9	11.2	12.3	12.2
	Vulnerability (biomass)	14.9	15.4	15.5	15.6	15.8
	Target (abundance)	64.5	77.2	71	94.6	76.5
	Target (biomass)	69.7	70	85.7	97.9	85
	Endemic (abundance)	48.9	69.3	52	64.6	55
	Endemic (biomass)	234.8	245.3	250.3	252.7	313.4

doi:10.1371/journal.pone.0155634.t002

Multivariate regression trees

Multivariate regression trees based on either the multibeam or predicted habitats, with the direct observer habitats and depth, were similar, explaining 25% and 31% of variance in assemblage composition, respectively (Fig 4). In both cases, the abiotic variable from the direct observer habitats was associated with the first split, explaining 38% of the variance in assemblage composition by differentiating sandy (16%) from reef habitats (22%). Depth further differentiated shallow (3%) from deeper habitats (3%). Only the last split differed between models, based on either the standard deviation in elevation within a 50-km radius from the multibeam variables (Fig 4A) or algal diversity from predicted habitats (Fig 4B). This last split had little relevance in the first (multibeam) case, only differentiating three samples characterized by a higher abundance of zooplanktivores (*Chromis westaustralis*). By contrast, in the second (predicted habitats) case, the last split led to a more even partitioning of samples, with the last cluster being characterized by medium to high vulnerability omnivores (*Parma mccullochi*), herbivores (*Meuschenia hippocrepis*), zooplanktivores (*Pseudocaranx* sp.) and

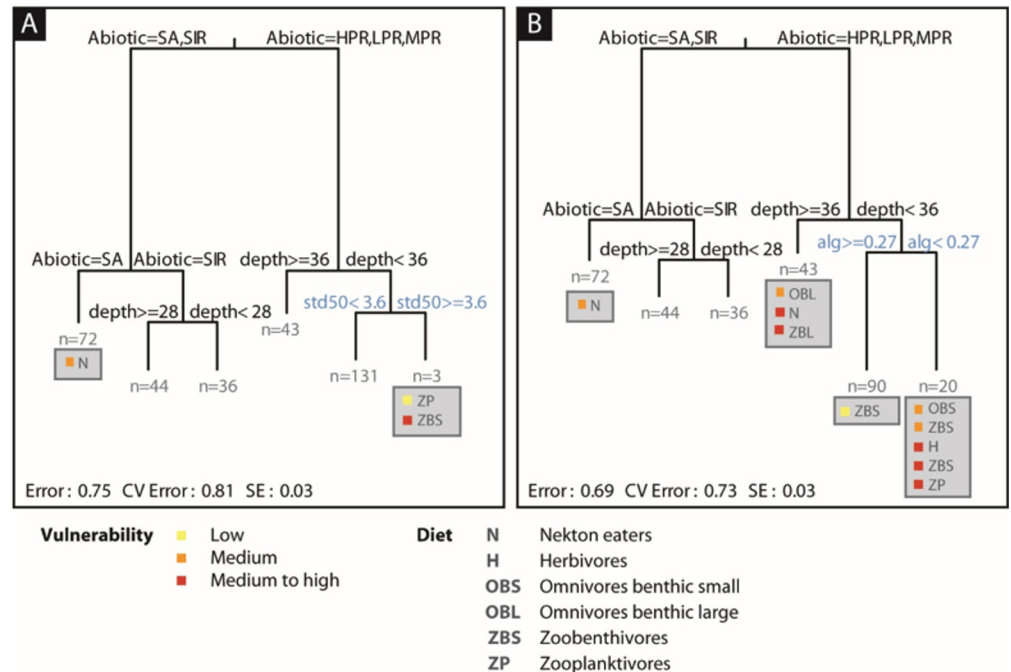


Fig 4. Multivariate regression tress (MRT) of fish functional groups based on A) multibeam data and B) predicted habitats data. Both MRTs also contained depth plus abiotic and biotic predictors from direct observer habitat classification. Multibeam and predicted habitat data were numeric; the abiotic and biotic predictor variables were categorical (five abiotic and six biotic classes). SA = sand, SIR = sand inundated reef, HPR/LPR/MPR = high/medium/low profile reef.

doi:10.1371/journal.pone.0155634.g004

zoobenthivores (*Coris auricularis*, *Notolabrus parilus*, *Pseudolabrus biserialis*, *Ophtalmolepis lineolata*, *Epinephelides armatus*, *Thalassoma lunare*).

The same models without the direct observer habitats explained only 3% of variance in the first case (i.e. multibeam; CVRE = 1.01), but 24% in the second case (i.e. predicted habitats; CVRE = 0.98). The latter tree consisted of a single split (general biota, *gen_bio*). Multivariate regression trees of fish biomass performed poorly with only 12–14% variance explained depending on the predictors used (RE 0.86–0.88; CVRE 0.88–1.01).

Discussion

Better predictive models of biodiversity are needed to support conservation and management of marine ecosystems under increasing anthropogenic pressures. Here we explored the use of habitat data obtained by three different classification methods for the development of models predicting a range of both uni- and multivariate fish biodiversity metrics. While in most cases, reasonable models could be developed with remotely sensed data, models that incorporated direct observer habitats consistently outperformed those that did not. Models for abundance also consistently outperformed corresponding models for biomass, and metrics of endemism were particularly difficult to model.

Remote sensing methods, such as multibeam, can provide valuable data for predicting biodiversity metrics [9,10]. Here, models developed using multibeam habitats alone were able to explain a substantial proportion of CV deviance for fish total species richness (56%) and total abundance (39%). Other studies predicting similar metrics from remotely sensed data have obtained similar or even better performance [16,33,34]. As such, our results reaffirm that

predictors generated by remote sensing can offer substantial utility in conservation planning and resource management. However, our results also illustrate limitations in the ability of remotely sensed data to predict biodiversity metrics.

Whilst remotely sensed data have been used to model fish species richness, abundance, and biomass (e.g., [33]), such models have not previously been compared to models generated using alternative predictors. Here, when we did so, models generated with multibeam habitats alone were consistently outperformed by those that included direct observer habitats. The abiotic variable from the direct observer habitats was also the single most influential predictor, possibly because multibeam data does not always enable differentiation between the presence or absence of certain habitats. For example, no detectable difference in multibeam imagery was found before and after the experimental removal of 100m² of kelp from three separate sites [15]. Studies have also shown that seagrass beds can be difficult to detect based on multibeam imagery, particularly when plants are present at low densities [35,36] or when plants are associated with certain substrate types [37]. Thus, there may be ecologically important aspects of the environment that are not captured well in multibeam data, but that are captured by direct observer habitat classification.

Consequently, modelling biodiversity with remotely sensed data represents a compromise between model performance and data availability. While predictors generated from direct observer habitat classification may provide valuable additional information when modelling fish biodiversity metrics, they can be prohibitively expensive to gather at sufficient resolution over large areas like Western Australia. Indeed, estimating biodiversity features over relatively large scales is a key problem [16,38]. Over these large scales remote sensing may offer the only cost-feasible way of generating high-resolution, continuous spatial data that can be used to generate predictions of biodiversity features, such as habitat classes or community level metrics.

Producing predicted habitats requires an additional resource investment, beyond the collection of remotely sensed data. Here, this greater investment did not necessarily result in better predictors for modelling fish biodiversity metrics. For five of the univariate metrics, models developed using predicted habitats resulted in higher CV deviance explained than those developed using multibeam habitats. However, for the other 4 univariate metrics we observed the opposite and in both cases the difference in CV deviance explained was small. BRT models built using multibeam habitats also had consistently lower mean prediction error than those developed with predicted habitats, which suggests that when predicting univariate metrics such as species richness over broad spatial scales, predictors from multibeam data may be more robust. Using multibeam data directly also avoids issues involved in the development of habitat classification models, such as the lack of a standardised method of classification, variable classification accuracy and variable amounts of ground-truthing [10]. Habitat classification maps based on remote sensing are an important part of many marine spatial conservation planning processes [39–42]. However, if habitat maps are not already available for a particular area of interest, our results suggest there may be little to no value in investing additional resources to develop them for the purpose of providing predictors for modelling univariate metrics of biodiversity. There is some indication here that predicted habitat classifications may provide better predictors when developing multivariate models, though more research is needed to understand this potential.

Some biodiversity metrics were more difficult to predict than others. Here, biomass was more difficult to predict than abundance, with lower CV deviance explained and higher mean prediction error. Because biomass is a function of abundance and individual length, estimates of biomass are likely to be more uncertain than non-composite metrics because of compounding errors [43]. Whilst all of the metrics explored here could provide valuable information for conservation planning, our results suggest that some, such as those for endemic species, may not be tractable with predictors from habitat data alone. Modelling others, such as abundance

of fisheries targeted species, may require more habitat data than can be generated from remote sensing or predicted habitats alone, and thus, would have limited applicability. Nonetheless, we were able to generate reasonable predictive models of the abundance of vulnerable fishes using only multibeam or predicted habitat data. Having a metric of fish vulnerability provides an important planning tool for both conservationists and fisheries managers [17]. Modelling fish vulnerability over areas such as Western Australia would provide valuable data on where to prioritise effort for both maintaining fish biodiversity and fish stocks, and could inform the positioning of spatial management units such as protected areas or localised fishing-gear modifications.

Biodiversity metrics are impacted by factors other than environmental conditions, and fish metrics will very likely be impacted by fishing. Through the selective extraction of targeted species, fishing alters the composition of communities such that the remaining fishes do not fully reflect the original community composition that would otherwise be determined by local biotic and abiotic conditions. Fishing impacts are very likely to have been present at Rottneest Island as it is a popular location for boat-based recreational fishing [44] and key target species are classified as overexploited in this area [45]. Because fishing generally targets larger individuals of higher trophic level species [46,47], fishing could be expected to have a greater impact on measures of fish biomass than abundance, which may be an additional cause of the poorer performance of the biomass models observed here. Indeed, incorporating fishing pressure, or spatial closures to fishing, alongside environmental predictors can improve model performance when predicting community assemblages [48] and classifying benthic biotopes [49]. Therefore, exploration of ways to incorporate fishing into predictive models of fish biodiversity metrics, over broad spatial scales, would likely be valuable. However, data on the spatial distribution of fishing effort often does not exist [50] or is not accessible [51]. Thus, future studies exploring the incorporation of historical and/or current fishing pressure may need to investigate the use of surrogates for fishing.

Individual habitat predictors had varying influence on the models of fish biodiversity metrics tested here, reflecting how the different metrics respond to different environmental drivers. Fish species richness can be influenced by habitat complexity (e.g. [52]), with increasing complexity leading to increased richness. In general sand patches are species poor, reefs are species rich, and significantly more species are found on high profile rather than low profile reefs [53,54]. Here we found that species richness was best predicted by the abiotic variable from the direct observer habitats, which consisted of five classes that relate very closely to habitat complexity: high/medium/low profile reef, sand inundated reef, sand. Furthermore, the second most influential predictor for species richness was 'range', a multibeam variable that describes the topographic variation in an area. The abiotic variable from the direct observer habitats and/or the variable range from the multibeam habitats were also the most influential predictors for the abundances and biomasses of fisheries targeted species, possibly reflecting the association of many of these species with higher complexity reefs. For endemic species (abundance and biomass) the relative influence of either the abiotic variable from the direct observer habitats or range from multibeam habitats was minimal. However, the predicted habitat class 'reef' (probability of occurrence) was the most influential of the non-multibeam predictors. This suggests that whilst many endemic species may be reef associated, they are not, as a group, associated with a particular level of reef profile/habitat complexity. Improved understanding of the relative influence of environmental drivers of a broader set of biodiversity metrics, such as those modelled here, could contribute to improved spatial conservation planning processes.

BRT and MRT are emerging techniques for modelling nonlinear species-habitat relationships. Both methods can accommodate different predictor types. Here, for example, we used both categorical and continuous data. These methods can also accommodate missing data, fit

complex nonlinear relationships and model interaction effects between predictors [26,27]. When modelling fish species richness, BRT has outperformed more established methods such as multiple linear regression, general adaptive models and neural networks on multiple occasions [16,38,55,56]. Both BRT and MRT can be used to predict across sites for which only environmental data are available [26,27]. Thus BRTs and MRTs could be used to generate maps of predicted biodiversity metrics from remotely sensed data over large areas, which could complement predicted habitat maps and aid conservation planning and natural resource management. Here we demonstrate that at least three community level metrics (total species richness, total abundance and abundance of vulnerable species) can be usefully predicted using BRT and remotely sensed data alone, and as such, how BRT models could inform conservation planning and natural resource management over large, data limited areas, such as the coastal seas off Western Australia. Ultimately, the utility of using remotely sensed data in BRT and MRT models to generate predictions of biodiversity metrics over large areas will depend on the transferability of models generated with sample data collected over a limited number of sites to unsampled areas [57]. Future studies now need to assess how transferable such models are, and investigate, the factors that impact model transferability.

Conclusion

Improved understanding of the distribution of a wide range of biodiversity metrics would have great utility for conservation and management, especially if that understanding was based on remotely sensed data. However, it seems that less standard metrics of biodiversity, such as vulnerability to exploitation or endemism, will be difficult to model. Moreover, whilst remotely sensed data can adequately predict total species richness, total abundance, and total biomass, predictive performance of such models can be improved by the addition of data from direct observer habitat classification. As such, it appears that the drivers of some metrics of fish biodiversity are not captured by habitat classification, and that there are aspects of habitats that are important for predicting fish biodiversity that are not readily captured through the use of remotely sensed data alone.

Supporting Information

S1 File. Supplementary materials: additional methods and results.
(DOCX)

S1 Table. Descriptions of all potential predictor variables.
(DOCX)

S2 Table. Parameters for 'best' BRT models.
(DOCX)

S3 Table. Ranges of CV Deviance and MPE for 'best' BRT models.
(DOCX)

S4 Table. Selected predictor variables.
(DOCX)

Acknowledgments

This work was done under the auspices of the Marine Biodiversity Hub. Our thanks to Dr K van Niel, for her contribution to the Marine Futures Project, and to three anonymous reviewers for their constructive and insightful comments.

Author Contributions

Conceived and designed the experiments: KLY CM MJC JJM. Performed the experiments: BTR JJM. Analyzed the data: KLY CM JJM. Wrote the paper: KLY CM MJC JJM BTR. Prepared figures: KLY CM BTR.

References

1. Beaumont NJ, Austen MC, Atkins JP, Burdon D, Degraer S, Dentinho TP, et al. Identification, definition and quantification of goods and services provided by marine biodiversity: implications for the ecosystem approach. *Mar Pollut Bull.* 2007; 54: 253–65. doi: [10.1016/j.marpolbul.2006.12.003](https://doi.org/10.1016/j.marpolbul.2006.12.003) PMID: [17266994](https://pubmed.ncbi.nlm.nih.gov/17266994/)
2. Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, et al. Impacts of biodiversity loss on ocean ecosystem services. *Science (80-)*. 2006; 314: 787–790.
3. Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, Katona SK, et al. An index to assess the health and benefits of the global ocean. *Nature.* Nature Publishing Group; 2012; 488: 615–20. doi: [10.1038/nature11397](https://doi.org/10.1038/nature11397)
4. Halpern BS, Walbridge S, Selkoe KA, Kappel C V, Micheli F, D'Agrosa C, et al. A global map of human impact on marine ecosystems. *Science (80-)*. 2008; 319: 948–52. doi: [10.1126/science.1149345](https://doi.org/10.1126/science.1149345)
5. Fisher R, Radford BT, Knowlton N, Brainard RE, Michaelis FB, Caley MJ. Global mismatch between research effort and conservation needs of tropical coral reefs. *Conserv Lett.* 2011; 4: 64–72. doi: [10.1111/j.1755-263X.2010.00146.x](https://doi.org/10.1111/j.1755-263X.2010.00146.x)
6. Fisher R, Knowlton N, Brainard RE, Caley MJ. Differences among major taxa in the extent of ecological knowledge across four major ecosystems. *PLOS One.* 2011; 6: e26556. doi: [10.1371/journal.pone.0026556](https://doi.org/10.1371/journal.pone.0026556) PMID: [22073172](https://pubmed.ncbi.nlm.nih.gov/22073172/)
7. Mellin C, Delean S, Caley J, Edgar G, Meekan M, Pitcher R, et al. Effectiveness of biological surrogates for predicting patterns of marine biodiversity: a global meta-analysis. *PLOS One.* 2011; 6: e20141. doi: [10.1371/journal.pone.0020141](https://doi.org/10.1371/journal.pone.0020141) PMID: [21695119](https://pubmed.ncbi.nlm.nih.gov/21695119/)
8. Sutcliffe PR, Pitcher CR, Caley MJ, Possingham HP. Biological surrogacy in tropical seabed assemblages fails. *Ecol Appl.* 2012; 22: 1762–71. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23092013> PMID: [23092013](https://pubmed.ncbi.nlm.nih.gov/23092013/)
9. Mellin C, Andréfouët S, Kulbicki M, Dalleau M, Vigliola L. Remote sensing and fish-habitat relationships in coral reef ecosystems: review and pathways for multi-scale hierarchical research. *Mar Pollut Bull.* Elsevier Ltd; 2009; 58: 11–9. doi: [10.1016/j.marpolbul.2008.10.010](https://doi.org/10.1016/j.marpolbul.2008.10.010)
10. Brown CJ, Smith SJ, Lawton P, Anderson JT. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuar Coast Shelf Sci.* Elsevier Ltd; 2011; 92: 502–520. doi: [10.1016/j.ecss.2011.02.007](https://doi.org/10.1016/j.ecss.2011.02.007)
11. Langlois TJ, Radford BT, Van Niel KP, Meeuwig JJ, Pearce AF, Rousseaux CSG, et al. Consistent abundance distributions of marine fishes in an old, climatically buffered, infertile seascape. *Glob Ecol Biogeogr.* 2012; 21: 886–897. doi: [10.1111/j.1466-8238.2011.00734.x](https://doi.org/10.1111/j.1466-8238.2011.00734.x)
12. Letessier TB, Meeuwig JJ, Gollock M, Groves L, Bouchet PJ, Chapuis L, et al. Assessing pelagic fish populations: The application of demersal video techniques to the mid-water environment. *Methods Oceanogr.* Elsevier Ltd; 2013; 8: 41–55. doi: [10.1016/j.mio.2013.11.003](https://doi.org/10.1016/j.mio.2013.11.003)
13. Watson D, Harvey E, Fitzpatrick B. Assessing reef fish assemblage structure: how do different stereo-video techniques compare? *Mar Biol.* 2010; 1–31. Available: <http://link.springer.com/article/10.1007/s00227-010-1404-x>
14. Huang Z, Nichol SL, Siwabessy JPW, Daniell J, Brooke BP. Predictive modelling of seabed sediment parameters using multibeam acoustic data: a case study on the Carnarvon Shelf, Western Australia. *Int J Geogr Inf Sci.* 2012; 26: 283–307. doi: [10.1080/13658816.2011.590139](https://doi.org/10.1080/13658816.2011.590139)
15. van Rein H, Brown CJ, Quinn R, Breen J, Schoeman D. An evaluation of acoustic seabed classification techniques for marine biotope monitoring over broad-scales (>1 km²) and meso-scales (10 m²–1 km²). *Estuar Coast Shelf Sci.* 2011; 93: 336–349. doi: [10.1016/j.ecss.2011.04.011](https://doi.org/10.1016/j.ecss.2011.04.011)
16. Pittman SJ, Christensen JD, Caldow C, Menza C, Monaco ME. Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecol Modell.* 2007; 204: 9–21. doi: [10.1016/j.ecolmodel.2006.12.017](https://doi.org/10.1016/j.ecolmodel.2006.12.017)
17. Cheung W, Watson R, Morato T, Pitcher T, Pauly D. Intrinsic vulnerability in the global fish catch. *Mar Ecol Prog Ser.* 2007; 333: 1–12. doi: [10.3354/meps333001](https://doi.org/10.3354/meps333001)
18. Hutchins J, Pearce A. Influence of the Leeuwin current on recruitment of tropical reef fishes at Rottnest Island, Western Australia. *Bull Mar Sci.* 1994; 54: 245–255.

19. GWA. Rottnest Island Management Plan 2009–2014. Revitalised and moving forward. The Government of Western Australia. 2014.
20. Radford B, Van Niel KP, Holmes K. WA Marine Futures: Benthic modelling and mapping final report. The University of Western Australia. 2008.
21. Stevens DL, Olsen AR, Althaus F, Hill N, Ferrari R, Edwards L, et al. A Standardised Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery : The CATAMI Classification Scheme. 2003; 610: 1–18. doi: [10.1371/journal.pone.0141039](https://doi.org/10.1371/journal.pone.0141039)
22. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC press. 1984.
23. De'ath G, Fabricius K. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*. 2000; 81: 3178–3192. Available: [http://www.esajournals.org/doi/abs/10.1890/0012-9658\(2000\)081%5B3178:CARTAP%5D2.0.CO%3B2](http://www.esajournals.org/doi/abs/10.1890/0012-9658(2000)081%5B3178:CARTAP%5D2.0.CO%3B2)
24. SeaGIS. PhotoMeasure and EventMeasure. SeaGIS Pty Ltd. www.seagis.com.au. PhotoMeasure and EventMeasure. SeaGIS Pty Ltd. www.seagis.com.au. 2008.
25. Froese R, Pauly D. FishBase. Available: <http://www.fishbase.org>, version (08/2014). In: (Editors) FishBase. www.fishbase.org, version (08/2014). 2008.
26. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. 2008; 77: 802–13. doi: [10.1111/j.1365-2656.2008.01390.x](https://doi.org/10.1111/j.1365-2656.2008.01390.x) PMID: [18397250](https://pubmed.ncbi.nlm.nih.gov/18397250/)
27. De'Ath G. Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology*. 2002; 83: 1105–1117. Available: [http://www.esajournals.org/doi/abs/10.1890/0012-9658\(2002\)083%5B1105:MRTANT%5D2.0.CO%3B2](http://www.esajournals.org/doi/abs/10.1890/0012-9658(2002)083%5B1105:MRTANT%5D2.0.CO%3B2)
28. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop)*. 2013; 36: 27–46. doi: [10.1111/j.1600-0587.2012.07348.x](https://doi.org/10.1111/j.1600-0587.2012.07348.x)
29. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available: <http://www.R-project.org/>. 2013.
30. Sutcliffe PR, Mellin C, Pitcher CR, Possingham HP, Caley MJ. Regional-scale patterns and predictors of species richness and abundance across twelve major tropical inter-reef taxa. *Ecography (Cop)*. 2014; 37: 162–171. doi: [10.1111/j.1600-0587.2013.00102.x](https://doi.org/10.1111/j.1600-0587.2013.00102.x)
31. De'Ath G. Boosted trees for ecological modeling and prediction. *Ecology*. 2007; 88: 243–51. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17489472> PMID: [17489472](https://pubmed.ncbi.nlm.nih.gov/17489472/)
32. Dufrene M, Legendre P. Species Assemblages and Indicator Species : the Need for a Flexible Asymmetrical Approach. *Ecol Monogr*. 1997; 67: 345–366.
33. Pittman SJ, Costa BM, Battista TA. Using Lidar Bathymetry and Boosted Regression Trees to Predict the Diversity and Abundance of Fish and Corals. *J Coast Res*. 2009; 10053: 27–38. doi: [10.2112/SI53-004.1](https://doi.org/10.2112/SI53-004.1)
34. Mellin C, Parrott L, Andréfouët S, Bradshaw CJA, MacNeil MA, Caley MJ. Multi-scale marine biodiversity patterns inferred efficiently from habitat image processing. *Ecol Appl*. 2012; 22: 792–803. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22645811> PMID: [22645811](https://pubmed.ncbi.nlm.nih.gov/22645811/)
35. Davidson DM, Hughes DJ. Zostera Biotopes (Volume I). An Overview of Dynamics and Sensitivity Characteristics for Conservation Management of Marine SACs. UK Marine SACs Project, Scottish Association for Marine Science, Dunstaffnage. 1998.
36. Mulhearn PJ. Mapping seabed vegetation with sidescan sonar. Defence Science & Technology Organisation. DSTO-TN-0381. Aeronautical and Maritime Research Laboratory. Victoria, Australia. 2001.
37. De Falco G, Tonielli R, Di Martino G, Innangi S, Simeone S, Michael Parnum I. Relationships between multibeam backscatter, sediment grain size and Posidonia oceanica seagrass distribution. *Cont Shelf Res*. Elsevier; 2010; 30: 1941–1950. doi: [10.1016/j.csr.2010.09.006](https://doi.org/10.1016/j.csr.2010.09.006)
38. Knudby A, LeDrew E, Brenning A. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. *Remote Sens Environ*. Elsevier Inc.; 2010; 114: 1230–1241. doi: [10.1016/j.rse.2010.01.007](https://doi.org/10.1016/j.rse.2010.01.007)
39. Margules CR, Pressey RL. Systematic conservation planning. *Nature*. 2000; 405: 243–253. PMID: [10821285](https://pubmed.ncbi.nlm.nih.gov/10821285/)
40. Yates KL, Schoeman DS. Incorporating the spatial access priorities of fishers into strategic conservation planning and marine protected area design: reducing cost and increasing transparency. *ICES J Mar Sci*. 2014; doi: [10.1093/icesjms/fsu122](https://doi.org/10.1093/icesjms/fsu122)
41. Giakoumi S, Grantham HS, Kokkoris GD, Possingham HP. Designing a network of marine reserves in the Mediterranean Sea with limited socio-economic data. *Biol Conserv*. Elsevier Ltd; 2011; 144: 753–763. doi: [10.1016/j.biocon.2010.11.006](https://doi.org/10.1016/j.biocon.2010.11.006)

42. Yates KL, Schoeman DS, Klein CJ. Ocean zoning for conservation, fisheries and marine renewable energy: assessing trade-offs and co-location opportunities. *J Environ Manage.* 2015; 152: 201–209. doi: [10.1016/j.jenvman.2015.01.045](https://doi.org/10.1016/j.jenvman.2015.01.045) PMID: [25684567](https://pubmed.ncbi.nlm.nih.gov/25684567/)
43. Kimmerera W, Aventa S, Bollensa Teven M, Feyrerb F, Grimaldob LF, Moylec PB, et al. Variability in Length–Weight Relationships Used to Estimate Biomass of Estuarine Fish from Survey Data. *Trans Am Fish Soc.* 2005; 134.
44. Sumner NR, Williamson PC, Blight SJ, Gaughan DJ. A 12-month survey of recreational boat-based fishing between Augusta and Kalbarri on the West Coast of Western Australia during 2005–06. Fisheries research report no. 177. Fisheries Research Division. Western Australian Fisheries and Marine Research Labor. 2008.
45. Wise BS, St John J, Lenanton RC. Spatial scales of exploitation among populations of demersal scalefish: implications for management. Part 1: Stock status of the key indicator species for the demersal scalefishery in the Western Coast Bioregion. Final FRDC Report—Project 2003 / 052. *Fish.* 2007.
46. Myers RA, Worm B. Rapid worldwide depletion of predatory fish communities. *Nature.* 2003; 423: 280–3. doi: [10.1038/nature01610](https://doi.org/10.1038/nature01610) PMID: [12748640](https://pubmed.ncbi.nlm.nih.gov/12748640/)
47. Pauly D, Watson R, Alder J. Global trends in world fisheries: impacts on marine ecosystems and food security. *Philos Trans R Soc Lond B Biol Sci.* 2005; 360: 5–12. doi: [10.1098/rstb.2004.1574](https://doi.org/10.1098/rstb.2004.1574) PMID: [15713585](https://pubmed.ncbi.nlm.nih.gov/15713585/)
48. Jimenez H, Dumas P, Ponton D, Ferraris J. Predicting invertebrate assemblage composition from harvesting pressure and environmental characteristics on tropical reef flats. *Coral Reefs.* 2011; 31: 89–100. doi: [10.1007/s00338-011-0820-9](https://doi.org/10.1007/s00338-011-0820-9)
49. Richmond S, Stevens T. Classifying benthic biotopes on sub-tropical continental shelf reefs: How useful are abiotic surrogates? *Estuar Coast Shelf Sci.* Elsevier Ltd; 2014; 138: 79–89. doi: [10.1016/j.ecss.2013.12.012](https://doi.org/10.1016/j.ecss.2013.12.012)
50. Yates KL, Schoeman DS. Spatial Access Priority Mapping (SAPM) with Fishers: A Quantitative GIS Method for Participatory Planning. *PLOS One.* Goldcoast, Queensland, Australia; 2013; 8: e68424. doi: [10.1371/journal.pone.0068424](https://doi.org/10.1371/journal.pone.0068424)
51. Hinz H, Murray LG, Lambert GI, Hiddink JG, Kaiser MJ. Confidentiality over fishing effort data threatens science and management progress. *Fish Fish.* 2013; 14: 110–117. doi: [10.1111/j.1467-2979.2012.00475.x](https://doi.org/10.1111/j.1467-2979.2012.00475.x)
52. Gratwicke B, Speight MR. The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *J Fish Biol.* 2005; 66: 650–667.
53. Molles MC. Fish species Diversity on Model and Natural Reef Patches : Experimental Insular Biogeography. *Ecol Monogr.* 1978; 48.
54. Harman N, Harvey ES, Kendrick GA. Differences in fish assemblages from different reef habitats at Hamelin Bay, south-western Australia. *Mar Freshw Res.* 2003; 54: 177–184.
55. Ferrier S, Guisan A, Elith J, Graham CH, Anderson RP, Dudi M, et al. Novel methods improve prediction of species ' distributions from occurrence data. *Ecography (Cop).* 2006; 2: 121–151.
56. Leathwick J, Moilanen A, Francis M, Elith J, Taylor P, Julian K, et al. Novel methods for the design and evaluation of marine protected areas in offshore waters. *Conserv Lett.* 2008; 1: 91–102. doi: [10.1111/j.1755-263X.2008.00012.x](https://doi.org/10.1111/j.1755-263X.2008.00012.x)
57. Sequeira A, Mellin C, Lozano-Montes H, Vanderklift M, Babcock RC, Haywood M, et al. Transferability of predictive models of coral reef fish species richness. *J Appl Ecol.* 2016; In Press.