Vol. 420: 157–174, 2010
doi: 10.3354/meps08858

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published December 16

Habitat suitability for marine fishes using presence-only modelling and multibeam sonar

Jacquomo Monk[1], Daniel Ierodiaconou[1,*], Vincent L. Versace[1], Alecia Bellgrove[1], Euan Harvey[2], Alex Rattray[1], Laurie Laurenson[1], Gerry P. Quinn[1]

[1]School Life & Environmental Sciences, Faculty of Science and Technology, Deakin University, PO Box 423, Warrnambool, Victoria 3280, Australia

[2]School of Plant Biology (Oceans Institute M470), University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

ABSTRACT: Improved access to multibeam sonar and underwater video technology is enabling scientists to use spatially-explicit, predictive modelling to improve our understanding of marine ecosystems. With the growing number of modelling approaches available, knowledge of the relative performance of different models in the marine environment is required. Habitat suitability of 5 demersal fish taxa in Discovery Bay, south-east Australia, were modelled using 10 presence-only algorithms: BIOCLIM, DOMAIN, ENFA (distance geometric mean [GM], distance harmonic mean [HM], median [M], area-adjusted median [Ma], median + extremum [Me], area-adjusted median + extremum [Mae] and minimum distance [Min]), and MAXENT. Model performance was assessed using kappa and area under curve (AUC) of the receiver operator characteristic. The influence of spatial range (area of occupancy) and environmental niches (marginality and tolerance) on modelling performance were also tested. MAXENT generally performed best, followed by ENFA-GM and -HM, DOMAIN, BIOCLIM, ENFA-M, -Min, -Ma, -Mae and -Me algorithms. Fish with clearly definable niches (i.e. high marginality) were most accurately modelled. Generally, Euclidean distance to nearest reef, HSI-b (backscatter), rugosity and maximum curvature were the most important variables in determining suitable habitat for the 5 demersal fish taxa investigated. This comparative study encourages ongoing use of presence-only approaches, particularly MAXENT, in modelling suitable habitat for demersal marine fishes.

KEY WORDS: Species distribution modelling · Multibeam sonar · Towed-video · MAXENT · ENFA · BIOCLIM · DOMAIN

INTRODUCTION

Understanding species distribution and associated measures of habitat suitability are fundamental elements of any planning or management program (Franklin 1995, Pearce & Ferrier 2001). In the terrestrial context, models of species distribution have long been recognised as cost-effective and powerful tools to estimate species occurrence across a landscape where limited direct observations exist (Pearce & Ferrier 2001, Ferrier et al. 2002a,b, Zaniewski et al. 2002, Elith et al. 2006, Guisan et al. 2006, Hirzel et al. 2006, Braunisch et al. 2008). These models are predicated on the assumption that the spatial variation in environmental factors (e.g. topography) used by a species control (either directly or indirectly) its geographic distribution (Guisan & Zimmermann 2000). It is only recently that predictive species distribution models have been used to predict species occurrence in management of the marine environment. These include applications to identify high conservation priority areas (Cañadas et al. 2005), assess the spatial patterns of suitable habitat within protected areas (Monk et al. 2010), predict sites at risk of invasion by exotic species (Compton et al. 2010) and investigate the distribution of coral disease (Williams et al. 2010).

*Corresponding author.
Email: daniel.ierodiaconou@deakin.edu.au

Parallel to the development and application of species distribution modelling in the marine environment, is the increasing access to multibeam sonar (MBES) technology and underwater video systems. These technological developments, coupled with advances in geographic information systems and computational power, make it possible to survey large regions of seafloor with unprecedented accuracy and resolution (Nasby-Lucas et al. 2002, Iampietro et al. 2005, Wilson et al. 2007, Ierodiaconou et al. 2010). MBES datasets are ideal for the application of a variety of terrain-analysis techniques, which form predictor variable datasets for input into models (see Wilson et al. 2007).

While traditionally used for assessing sessile species, 'passive' underwater video systems such as drop video, towed/drift video and remotely operated vehicles (ROV) are increasingly being used as cost-effective, non-destructive methods for assessing marine fish species distributions (Morrison & Carbines 2006, Anderson & Yoklavich 2007). These video-based survey methods have significant advantages over traditional methods (e.g. SCUBA divers) in collecting fish occurrence data. They are capable of being deployed at depths and times that are dangerous for divers (Assis et al. 2007), provide a permanent record of survey (Watson et al. 2005), afford accurate positioning when coupled with differential GPS and acoustic positioning, enable high replication and, in the case of drift/towed and ROV, capture transitions between different habitat types (Spencer et al. 2005). However, video-based survey methods also have a number of disadvantages, including restricted field of view and the need for high water clarity (Murphy & Jenkins 2010). Mueller et al. (2006) estimated that a stationary underwater video camera was able to detect around 45 to 75% of large (i.e. 30 to 50 cm) trout in water with turbidity levels equivalent to those encountered in shallow temperate marine waters (i.e. 0 to 4 nephelometric turbity units, NTU). Additionally, towed video systems are constantly moving (i.e. up, down and side-to-side) and are often 'flown' over highly rugose reef systems covered in a dense canopy-forming macroalgae. These factors combined have the potential to underestimate species distributions because fish may simply be hiding under or camouflaged within the canopy and reef. Consequently, observations of demersal fishes derived from such towed video techniques may lead to the inclusion of false absences (i.e. a failure of the survey to detect the species when it is actually present). Incorporation of such 'false absences' has the potential to bias model predictions (Hirzel et al. 2001).

Integrating towed video data capture methods and MBES in a presence-only predictive modelling framework has the potential to better our understanding of the fine-scale spatial ecology and distribution of many marine fish species, while avoiding issues of false absences. Presence-only models have 2 unique attributes that make them potentially more useful than presence/absence models for estimating habitat suitability for marine demersal fishes based on towed video observations. First, presence-only models do not require the explicit constraints indicated by absence data. Considering the issues surrounding accurately surveying fish assemblages using towed video, it would be inappropriate to treat localities without an observed presence as unsuitable. Instead, presence-only approaches are based on constructing a model of a species' niche from locational records. This modelled niche/habitat signature can then be used to predict distribution, or suitability, within the available environment. Second, most presence-only models are designed to function well even when limited to very small occurrence datasets (Engler et al. 2004, Hernandez et al. 2006), meaning that useful models of habitat suitability can often be developed from very few presence locations (i.e. ~30 observations). This feature is particularly important because of the often small sample sizes obtained from marine fish surveys (Langlois et al. 2010).

There is a broad array of quantitative approaches available to model species habitat suitability based on presence-only datasets (e.g. Nix 1986, Stockwell & Noble 1992, Robertson et al. 2001, Hirzel et al. 2002, Phillips et al. 2006, Li & Hilbert 2008). Consequently, researchers are faced with the difficulty of selecting between numerous presence-only modelling approaches. Relatively few studies have compared more than 3 different methods on the same data, and the majority of these have focused on terrestrial, freshwater or simulated datasets (Elith & Burgman 2002, Ferrier et al. 2002b, Olden & Jackson 2002, Farber & Kadmon 2003, Loiselle et al. 2003, Brotons et al. 2004, Segurado & Araújo 2004, Elith et al. 2006, Tsoar et al. 2007). The effective application of presence-only modelling methods in the marine environment, however, has received less attention, specifically in relation to demersal fish species. We are aware of only 2 studies that have compared presence-only algorithms for marine organisms (MacLeod et al. 2008, Tittensor et al. 2009). MacLeod et al. (2008) used data on the occurrence of harbour porpoises to compare the predictive abilities of one presence/absence approach (generalised linear modelling) and 3 presence-only approaches (principal component analysis-based approach, genetic algorithm for rule-set prediction and ecological-niche factor analysis [ENFA]). They concluded that presence-only approaches can produce models of habitat suitability of marine species that are significantly better than random and exhibit comparable performances to presence/absence modelling approaches (MacLeod et al. 2008).

In the marine environment, studies have predominantly focused on how variations in predictor variables

(e.g. seafloor characteristics such as bathymetry or rugosity) affect the ability of the models to predict species distributions (e.g. Pittman et al. 2009). In the terrestrial context, in addition to the predictor variables, the influence of response variables (i.e. the ecological characteristics of the species occurrence data) has been more widely investigated (e.g. Araújo & Williams 2000, Pearce & Ferrier 2000, Manel et al. 2001, Pearce et al. 2001, Segurado & Araújo 2004, Sérgio et al. 2007). An organism with a narrow range of environmental requirements (i.e. clearly definable niche) is likely to return a better performing model in terms of model diagnostics (e.g. model performance measures such as kappa and Area Under Curve of the Receiver Operator Characteristic) (Pearce & Ferrier 2000, Brotons et al. 2004, Segurado & Araújo 2004, Elith et al. 2006). In contrast, it may be more difficult to define the habitat characteristics of an organism with a much wider range of environmental tolerances as there are more combinations of biophysical factors that could potentially be used as habitat.

In this study, we compared the relative performance of presence-only modelling approaches to predict habitat suitability of marine fish taxa using remotely-sensed occurrence and high-resolution MBES-derived seafloor datasets. Specifically, we used data on 5 demersal fish taxa and 2 measures of model performance to compare 10 different modelling algorithms: BIOCLIM, DOMAIN, ENFA (using the 7 available algorithms in the software) and MAXENT. Previous evidence indicates differences in model performance may depend on the species range and environmental niches (Pearce & Ferrier 2000, Brotons et al. 2004, Segurado & Araújo 2004, Elith et al. 2006). To investigate this, we also tested if variation in these factors influenced model performance. Finally, we identified key seafloor variables that influence the spatial distribution and discuss these in terms of the known ecology of each taxon investigated.

MATERIALS AND METHODS

Study site. The study site encompassed an area 42 km^2 that was situated on the western side of Cape Duquesne (38° 22′ S, 141° 21′ E) in Discovery Bay, south-eastern Australia (Fig. 1). The site ranged in

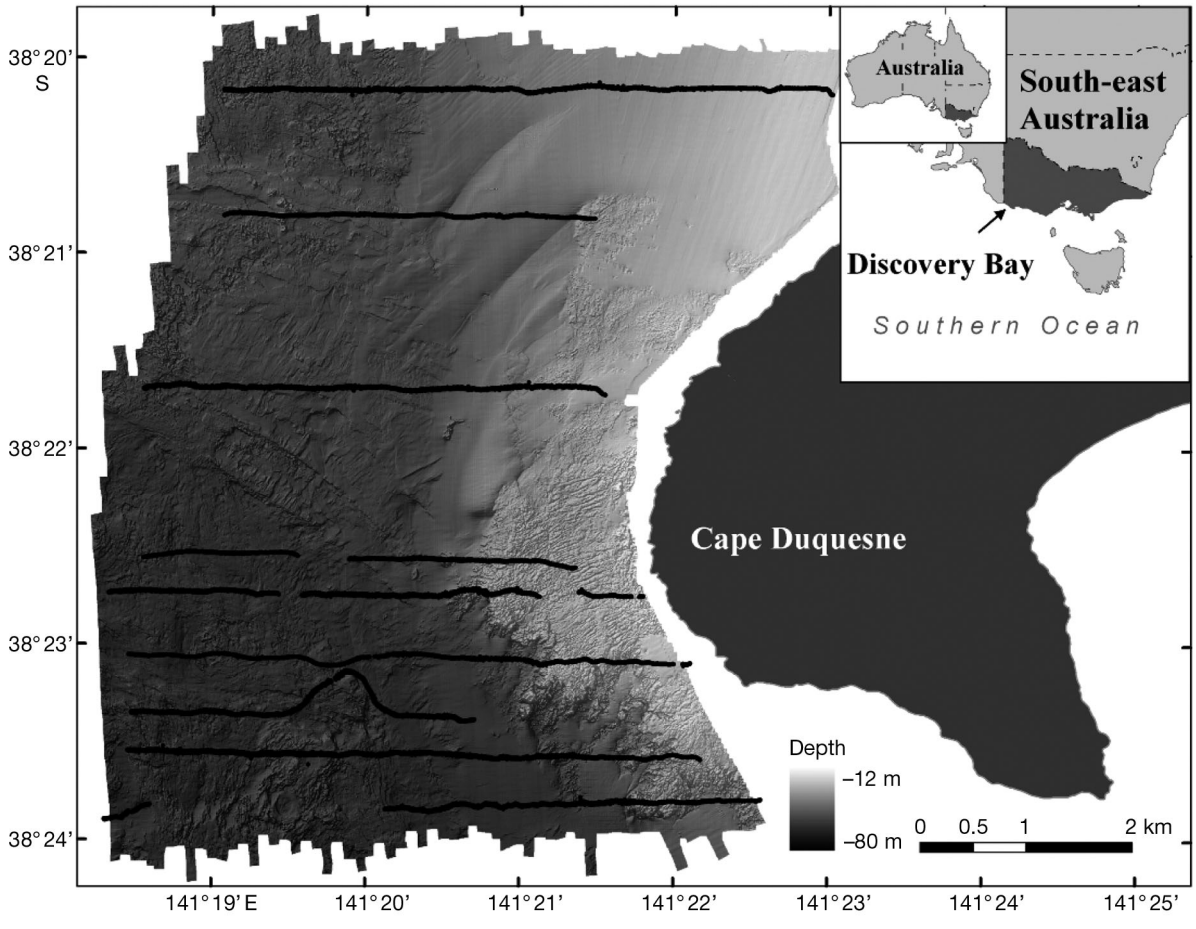

Fig. 1. Discovery Bay study area on the south-east Australian continental shelf. Shading: water depth. Black lines: towed video transects

depth from 12 to 80 m. Vertical basalt reef structures rise some 20 m from the seafloor, reflecting the region's dynamic volcanic history (Boutakoff 1963). The tops of these reef structures support diverse assemblages of red algae and kelps (dominated by *Ecklonia radiata, Phyllospora comosa* and *Durvillaea potatorum*), while the deeper regions are covered in sponges, ascidians, bryozoans and gorgonian corals (Ierodiaconou et al. 2007a).

Fish occurrence data for model training and testing. Towed-video transects, which were initially collected for benthic habitat mapping, were used to provide fish occurrence data for model training and testing. Nine transects, aligned perpendicular to the coast, were selected to encompass the main physical gradients (e.g. depth, topographic variation, exposure). These 9 transects covered 56 linear km of the study area (Fig. 1). Over 4 d (24, 25 March and 26, 27 April 2006) a micro remotely-operated vehicle (VideoRay Pro 3) was towed along the transects at 0.5 to 1 m s^{-1} (1 to 2 knots) to collect video data. The oblique angled camera was maintained ~2 m from the bottom using a vessel-mounted winch system. A text overlay containing a time stamp and transect ID were recorded with the video using a Sony MiniDV recorder. The video footage was interrogated to identify fish to the lowest possible taxonomic resolution. The exact spatial position (±5 m accuracy) of each fish taxon was then determined by matching the time stamp of the video with the corresponding survey positional data. The survey positional data was recorded through the integration of vessel location (Omnistar satellite dGPS), motion sensor (KVH) and acoustic camera positioning (Tracklink Ultra Short Baseline).

More than 7300 individual fish, representing 40 species, were observed. Some video frames included multiple individuals of the same species resulting in a total of 1648 observational events. From these observations, we used occurrence data for the 5 most commonly observed demersal fish taxa (Table 1). For each taxon, we made 10 random partitions of the occurrence localities. Each partition was created by randomly selecting 75% of the occurrence localities as training data, with the remaining 25% reserved for testing the resulting models (Table 1). We made 10 random partitions to assess the average performance of the algorithms, and to allow for statistical testing of observed differences in performance for each taxon.

Table 1. Partitioning of occurrence data into training and test localities

Taxon	Training	Testing	Total
Caesioperca spp.	710	237	946
Cheilodactylus nigripes	25	8	33
Notolabrus tetricus	47	16	62
Pempheris multiradiata	38	13	50
Pseudolabrus psittaculus	283	94	377

Spatial autocorrelation was investigated by examining Moran's *I* statistic on model residuals (i.e. for each species, the observed occurrence minus the predicted probability of occurrence given by each modelling approach). The distance classes and Moran's *I* statistics were computed using the freeware package SAM (Spatial Analysis in Macroecology). Only very weak global spatial autocorrelation (i.e. all species were <0.08) was found at the level of the first neighbour. Consequently, we assumed the partitioning of data into model training and testing datasets to be independent of each other.

MBES survey and derived seafloor characteristics used in model construction. Multibeam echosounder (MBES) derived bathymetry and backscatter variables were gridded to a 2.5 m^2 cell resolution. MBES data was acquired on 2 and 3 November (38 h) 2005 as part of the Victorian Marine Habitat Mapping Project (Ierodiaconou et al. 2007a). A detailed description of the MBES data acquisition is provided in Monk et al. (2010). Details on MBES processing methods are provided in Rattray et al. (2009). The processed bathymetry and backscatter were used to produce an additional 11 variables of the seafloor using a range of techniques and applications in ENVI 4.2 (RSI) and ArcGIS 9.3 (ESRI) based using 3 × 3 cell-analysis window (Table 2). To reduce the likelihood of model over-fitting, a covariance matrix and correlation tree, using a correlation coefficient threshold of 0.5 (Hirzel et al. 2002, Galparsoro et al. 2009), were used to reduce the 13 variables to the 8 least correlated (Table 3). Consequently, bathymetry, benthic position index (BPI), eastness, Euclidean distance to nearest reef, hue-saturation-intensity-blue (HSI-b; backscatter derivative), maximum curvature, northness and rugosity were retained for model construction (Fig. 2, Table 2). Euclidean distance to nearest reef was calculated using the spatial analyst tool in ArcGIS 9.3 based on an accurately modelled reef class (81% accuracy) extracted from a predicted substrata map of the study area. These variables were chosen as they captured the main seafloor characteristics and many of them are known to be important determinants of demersal fish habitat (e.g. distance to reef: Friedlander & Parrish 1998; bathymetric variance: Pittman et al. 2007; rugosity: Wedding et al. 2008). We limited the analysis to MBES derived seafloor characteristics because oceanographic variables are not available at the fine-scale resolution of the MBES datasets (2.5 m^2). In addition, because of the relatively small spatial scale of the study area, we expected that there would be little variation in oceanographic characteristics (e.g. temperature).

Table 2. Description of the 13 seafloor variables

Variable	Description	Software
Aspect (eastness, northness)	Aspect (azimuthal bearing of steepest slope) has an inherent circularity built in; to overcome this, 2 trigonometric transformations (Roberts 1986) were applied: northness (sin[aspect]) and eastness (cos[aspect]). These 2 variables represent proxies for exposure.	Spatial Analyst-ArcGIS 9.3
Bathymetry	Bathymetry provides a measure of water depth.	Fugro Starfix suite 9.1
Backscatter	Backscatter is the acoustic scattering of the seafloor that represents the softness and hardness of the substratum.	Fugro Starfix suite 9.1
Benthic position index	Measure of a location relative to the overall landscape. Calculated by comparing the elevation of a cell with the mean elevation of surrounding cells by the 3 analysis extents. Regions with positive values are higher than their surroundings, whereas areas with negative values are lower. Flatter areas have values closer to zero (Weiss 2001)	Benthic Terrain Modeller Tool for ArcGIS
Complexity	Complexity is the rate of change of the slope and is a measure of local variability in benthic terrain.	ENVI 4.2
Euclidean distance to nearest reef	Reef class from a substratum map, generated using a decision tree classifier (Ierodiaconou et al. 2007b) was used to calculate Euclidean distance (m) to nearest reef.	Spatial Analyst-ArcGIS 9.3
HSI (R, G, B)	Hue-saturation-intensity (HSI) is a transformation of backscatter (proxy for seafloor hardness/softness), initially developed to decrease noise in radar reflectance (Daily 1983). Since backscatter represents seafloor reflectance, a HSI transformation may improve the separation of high and low frequency signal-scattering properties of the substratum. The high and low frequency information is mapped to hue (chromatic) and intensity (achromatic) respectively, with a fixed saturation value. These HSI values are transformed into a red, green, blue (RGB) colour space to produce 3 variables.	ENVI 4.2
Maximum curvature	Maximum curvature provides the greatest curve of either the profile or plan convexity relative to the analysis window (Gallant & Wilson 1996).	ENVI 4.2
Rugosity	Rugosity provides the ratio of surface area to planar area within the analysis window and is to represent a measure of structural complexity (Lundblad et al. 2006).	Benthic Terrain Modeller Tool for ArcGIS
Slope	Slope is the maximum change in elevation between each cell and cells in its analysis neighbourhood (3 × 3 cells). Calculated in degrees from horizontal.	ENVI 4.2

Table 3. Correlation matrix used to assess independence between seafloor variables. A 0.5 threshold was applied to determine the least correlated variables that were retained for model construction (shown **in bold**). BPI: benthic position index; HSI: Hue-saturation-intensity (b: blue; g: green; r: red)

	Backscatter	Bathymetry	BPI	Complexity	Eastness	Euclidean distance to nearest reef	HSI-b	HSI-g	HSI-r	Maximum curvature	Northness	Rugosity
Bathymetry	−0.894											
BPI	−0.022	0.043										
Complexity	0.223	−0.168	0.134									
Eastness	0.289	−0.279	0.06	0.296								
Euclidean distance to nearest reef	−0.015	−0.125	−0.138	−0.607	−0.262							
HSI-b	0.152	0.036	0.048	0.186	0.138	−0.363						
HSI-g	−0.769	0.749	0.043	−0.114	−0.208	−0.143	−0.205					
HSI-r	0.59	−0.622	−0.049	0.027	0.131	0.246	−0.319	−0.683				
Maximum curvature	−0.023	0.057	0.279	0.331	0.089	−0.254	0.089	0.038	−0.091			
Northness	−0.058	0.06	0.008	−0.034	−0.089	−0.025	−0.083	0.04	0.035	−0.014		
Rugosity	−0.038	0.086	0.128	0.341	0.071	−0.266	0.111	0.051	−0.127	0.461	−0.023	
Slope	0.115	−0.049	0.106	0.679	0.141	−0.515	0.221	−0.038	−0.094	0.345	−0.025	0.375

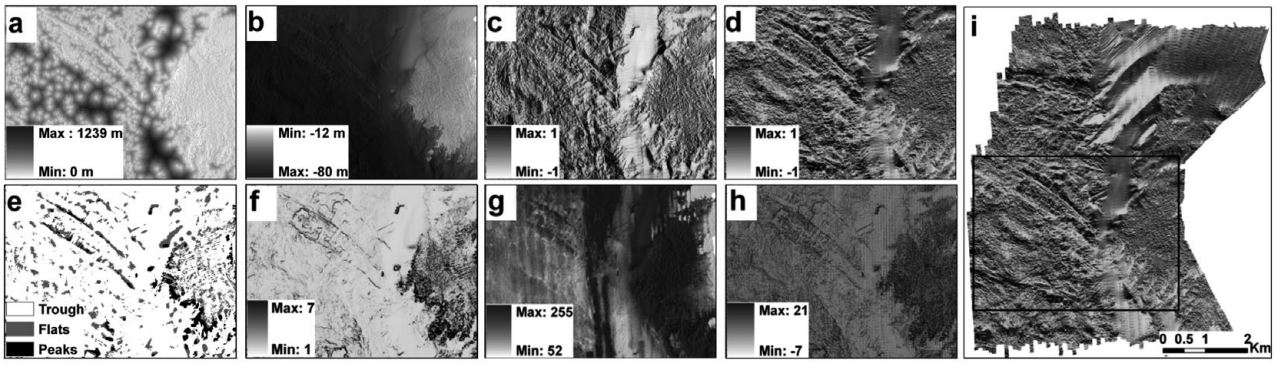

Fig. 2. Zoomed examples of the 8 variables used in model construction. Darker shading: high values. (a) Euclidean distance to nearest reef, (b) bathymetry, (c) eastness, (d) northness, (e) benthic position index, (f) rugosity, (g) hue-saturation-intensity-blue, (h) maximum curvature, (i) hillshade of study area showing extent of zoomed examples

Modelling approaches. Four distinct modelling approaches were used; however, ENFA outcomes were converted into geographic space using all 7 available algorithms (distance geometric mean [GM], distance harmonic mean [HM], median [M], area-adjusted median [Ma], median + extremum [Me], area-adjusted median + extremum [Mae] and minimum distance [Min]), resulting in 10 different models for each of the 5 fish taxa. Each algorithm was run on each of the 10 fish-occurrence data partitions, resulting in 500 model runs. Predictive models used were BIOCLIM (Busby 1986, Nix 1986), DOMAIN (Carpenter et al. 1993), ENFA (using all 7 available algorithms) (Hirzel et al. 2002, Braunisch et al. 2008), and MAXENT (Phillips et al. 2006). All of these methods are based on the concept of the ecological niche (Hutchinson 1957). Each method uses mathematical algorithms to define the ecological niche of the focal taxon based on the distribution of the occurrence records in multidimensional environmental space. Once this niche is defined, it is projected into geographic space to produce a predictive map of suitable habitat.

Both BIOCLIM and DOMAIN were run through DIVA-GIS (Hijmans et al. 2005, Hijmans & Graham 2006). BIOCLIM predicts suitable conditions in a 'bioclimatic envelope', consisting of a rectilinear region in environmental space representing the range (or some percentage thereof) of observed presence values in each environmental dimension. This envelope specifies the model in terms of percentiles or upper and lower tolerances of variables, and does not allow for regions of absence (i.e. 'holes') within the envelope. The concept is one of extremes and cores. In this study, we applied a percentile range of 95% (excluding 2.5% of the values either side of the rectilinear box in environmental space). Similarly, DOMAIN (Carpenter et al. 1993) uses the Gower metric, a distance measure that standardises each variable by its range over all occurrence observations to equalise the contribution of all variables. A predicted suitability index is given by computing the minimum distance in environmental space to any occurrence record.

ENFA was conducted through Biomapper 4 software (Hirzel et al. 2007). It is similar to principal components analysis, involving a linear transformation of the environmental space into orthogonal 'marginality' and 'specialisation' factors. Environmental suitability is then modelled by the 7 different distance-based algorithms in the transformed space. These algorithms make the assumptions that, on all ENFA factors, either the geometric mean, harmonic mean, median, or minimum distance of the species distribution offers the optimal habitat conditions, taking into account the density of observations in environmental space, with the following exceptions. The Ma algorithm assumes that, for all ENFA factors, the optimal habitat is the median of the habitat use to habitat availability ratio (Braunisch et al. 2008). The Me algorithm assumes that the optimal habitat suitability is, for the marginality factor, either the lowest or highest value, whereas for all specialisation factors, it is the median (as in the M algorithm) (Braunisch et al. 2008). Finally, the Mae algorithm is a combination of the Ma and Me algorithms. For all ENFA algorithms, across the 5 fish taxa, 3 factors were retained based on 'MacArthurs broken-stick' rule (MacArthur 1957), which compares the eigenvalue distribution of the factors to ensure that there is no overlap and that only those that are necessary are retained (i.e. with eigenvalues >1).

MAXENT (Phillips et al. 2006) uses the maximum entropy method for modelling species geographic distributions with presence-only data. MAXENT is a general-purpose, machine-learning method with a simple and precise mathematical formulation, and it has a number of aspects that make it well-suited for species distribution modelling (see Phillips et al. 2006). Models were trained using default settings; convergence threshold (0.00001), maximum iterations (1000), auto

features, regularisation multiplier (r = 1) and background points (10 000).

Model testing. Using the occurrence datasets that were set aside for model testing, model performance was evaluated using 2 methods: the threshold-independent AUC (area under the curve) of the ROC (receiver operating characteristic) (Fielding & Bell 1997) and threshold-dependent kappa statistic (Cohen 1960).

The ROC plots sensitivity (the fraction of occurrence records that are classified as presence) against 1 – specificity (the portion of absences points that are classified as absent) for all possible thresholds. A curve that maximises sensitivity for low values of the false-positive fraction is considered a good model and is quantified by calculating the AUC. An AUC value of 0.5 implies the model predicts species occurrence no better than random, and a value of 1.0 implies perfect prediction (Fielding & Bell 1997). While this technique was initially applied to presence/absence methods, it can be adapted to evaluate presence-only models (Wiley et al. 2003, Phillips et al. 2006). Plotting sensitivity against a random sample of background locations is sufficient to define an ROC curve (Wiley et al. 2003, Phillips et al. 2006). The ROC plot method has an advantage over confusion matrix-derived evaluation methods (for examples see Fielding & Bell 1997) because it does not require an arbitrary selection of a threshold above which prediction is considered positive: a procedure that can bias evaluations (Fielding & Bell 1997). A 1:1 ratio of presence:background point was used (i.e. if there were 237 occurrences, then 237 background points were randomly generated). For each individual taxon, these background points were randomly generated along transects where no fish taxa were observed. The AUC derived from the ROC plot of this study can be interpreted as a measure of the ability of the algorithm to discriminate between a suitable environmental condition and a random analysis pixel (background), rather than between suitable and unsuitable conditions, as an AUC developed with measured absences is interpreted (Phillips et al. 2006). The ROC curves and the AUC values were calculated in DIVA-GIS.

Although the continuous map of the probability of presence produced by distribution models is itself useful for many conservation applications (e.g. Wilson et al. 2005), it is often converted into a presence/absence map. Despite threshold-independent measures of model performance being widely preferred, such as the AUC, the reliability of these methods has recently been questioned (Lobo et al. 2008). Therefore, to enable a thorough evaluation of our models, we selected the threshold-dependent kappa statistic (Cohen 1960). The kappa statistic assesses the extent to which a model predicts occurrence at a rate higher than expected by chance. The results vary between 1.0 for perfect agreement and 0 for random agreement. We used the kappa-maximised threshold in DIVA-GIS. This threshold method calculates kappa scores for 100 threshold values (in 0.01 increments) and the one that provides maximum kappa is accepted as the threshold (Thuiller 2003). When multiple thresholds had the same kappa value, the mean threshold value was selected.

To compare the influence of the modelling algorithm on model performance (as measured by AUC and kappa), a permutational multivariate analysis of variance (PERMANOVA) was used (Anderson 2001). PERMANOVA was run with 999 permutations of the residuals under a reduced model. Post-hoc pairwise tests (*t*-tests) were conducted, again using 999 permutations. All analyses were tested at $\alpha = 0.05$. PERMANOVA was conducted using the PERMANOVA+ add-on in the statistical package PRIMER-E (Clarke & Gorley 2006).

Influence of species' spatial range and environmental niches on modelling performance. When examining model outputs of marine species distribution, much of the emphasis is on the influence of explanatory variables (e.g. Moore et al. 2009). However, the influence of species' spatial range and environmental niches is less often considered in the marine environment. To address this imbalance, a series of orthogonal contrasts were carried out on the modelled distributions of habitat suitability generated in this study. The rationale of these analyses was the assumption that the modelled distributions of habitat suitability of taxa with wider spatial range and environmental niches would be characterised by lower performance (in terms of kappa and AUC) relative to taxa that exhibit a narrower range and niche. The spatial range and environmental niches were determined using 3 measures: area of occupancy (AOO), marginality and tolerance. AOO is an estimate of geographic-range size that is actually occupied by a taxon (Gaston 1996). It was calculated by summing the number of 2.5-m^2 grid cells where a fish taxon was observed. Marginality is the difference between the species' optimum habitat and the mean environmental conditions in the study area. This is therefore representative of the species' ecological niche position. Tolerance describes the species' niche breadth by comparing the variability in the environmental conditions where the species occurs to the range of environmental conditions in the study area. These characteristics were calculated using BIOMAPPER (Hirzel et al. 2002).

Six sets of orthogonal contrasts (Field 2000) were used to test the influences of (1) marginality on mean kappa; (2) marginality on mean AUC; (3) tolerance on

mean kappa; (4) tolerance on mean AUC; (5) AOO on mean kappa; and (6) AOO on mean AUC. For the design of these contrasts, the 5 fish taxa were used to provide the measures of AOO, marginality and tolerance, while model diagnostics of the 10 algorithms were used to provide measures of mean kappa and mean AUC. Using a series of *t*-tests, orthogonal contrasts provide an efficient way to test specific hypothesis while maintaining strict control over the Type I error rate by reducing the amount of pairwise comparisons when compared to post-hoc testing (Field 2000). Homogeneity of variance in the dependent variables was assessed using Levene's test and normal distribution was assessed using Kolmogorov–Smirnov test (Field 2000).

Most important variables for characterisation of suitable habitat for demersal fishes. Using the best model run (i.e. MAXENT) for each fish taxon, we used a jackknife analysis of the regularised gain (a statistic that measures how well a variable distinguishes localities where taxa occur from the total area under study) with the training occurrence data. The variables with regularised gains ≥0.1 were considered important. To assess how the variation in these variables influenced suitable habitat, response curves (log contribution to prediction) were used. These curves show how prediction of suitable habitat changes as each seafloor variable is varied, keeping the remaining 7 variables at their average sample value. Predictor variable values with positive log contribution (i.e. >0) indicate higher habitat suitability.

RESULTS

Comparison of model performances

Generally, MAXENT was found to be the best performing model (Fig. 3). Subtle differences between each fish taxon, however, were observed (Fig. 3). Pairwise comparisons based on kappa and AUC indicated that MAXENT was significantly better than all other algorithms for *Notolabrus tetricus* and *Pseudolabrus psittaculus* (pairwise tests: $p < 0.05$). Similarly, pairwise comparisons based on kappa and AUC indicated that MAXENT was significantly better than all other algorithms with the exception of DOMAIN for *Caesioperca* spp. and *Cheilodactylus nigripes*. Results for *Pempheris multiradiata* were less defined. Pairwise comparisons based on AUC indicated that MAXENT was significantly better than ENFA Min (minimum distance), but was not significantly different to all other algorithms (Fig. 3h). Pairwise results from kappa indicated, however, that MAXENT was significantly better than all other algorithms with exception to ENFA GM (geometric mean) (Fig. 3g). In addition to differences in model performance, we observed considerable variation in the spatial distributions projected by the different algorithms (Fig. 4 shows examples of the highest and lowest performing algorithms for the 5 demersal fishes investigated).

Influence of species' spatial range and environmental niches on modelling performance

Orthogonal contrasts indicated that as marginality increased, there was a trend of increasing mean kappa and AUC (Table 4). There were no discernible trends when examining the influence of niche breadth (tolerance) or species spatial range (AOO) on mean AUC and mean kappa (Table 4).

Most important variables for characterisation of suitable habitat for demersal fishes

Variable importance varied for the 5 fish taxa based on the best performing MAXENT model run (Table 5). It is acknowledged that the fitted response curves (log contribution to prediction; Fig. 5) are graphical descriptions of how the different variables influence habitat suitability and do not describe the environmental limits of the taxa (Ysebaert et al. 2002).

For *Caesioperca* spp., the areas surrounding sheer drop-offs along the south-eastern region of the study area were predicted to be the most suitable habitat (Fig. 4b). Areas of fragmented, highly suitable habitat were also predicted throughout the deeper (>30 m) regions of the study area (Fig. 4b). The jackknife test of variable importance showed that bathymetry, Euclidean distance to nearest reef, HSI-b and rugosity were important for determining suitable habitat for *Caesioperca* spp. (Table 5). Response curves of these variables indicated that suitable habitat for *Caesioperca* spp. decreased with the increase in distance from reef (Fig. 5d). Similar patterns were observed for bathymetry and HSI-b variables (Fig. 5a,e). The response curves showed that as these values increased, so did habitat suitability of *Caesioperca* spp. However, the curves showed this trend up to a certain point (~32 m depth and ~80 HSI-b) and beyond that the habitat suitability plateaued. Variation in rugosity, however, had only a small influence on suitability, with values ~2 being most important (Fig. 5h).

Regions of highest habitat suitability for *Cheilodactylus nigripes* were predicted to be the deeper (>30 m) fringing reefs at the base of the drop-offs (Fig. 4d). The most important seafloor variables for defining suitable habitat were rugosity, Euclidean dis-

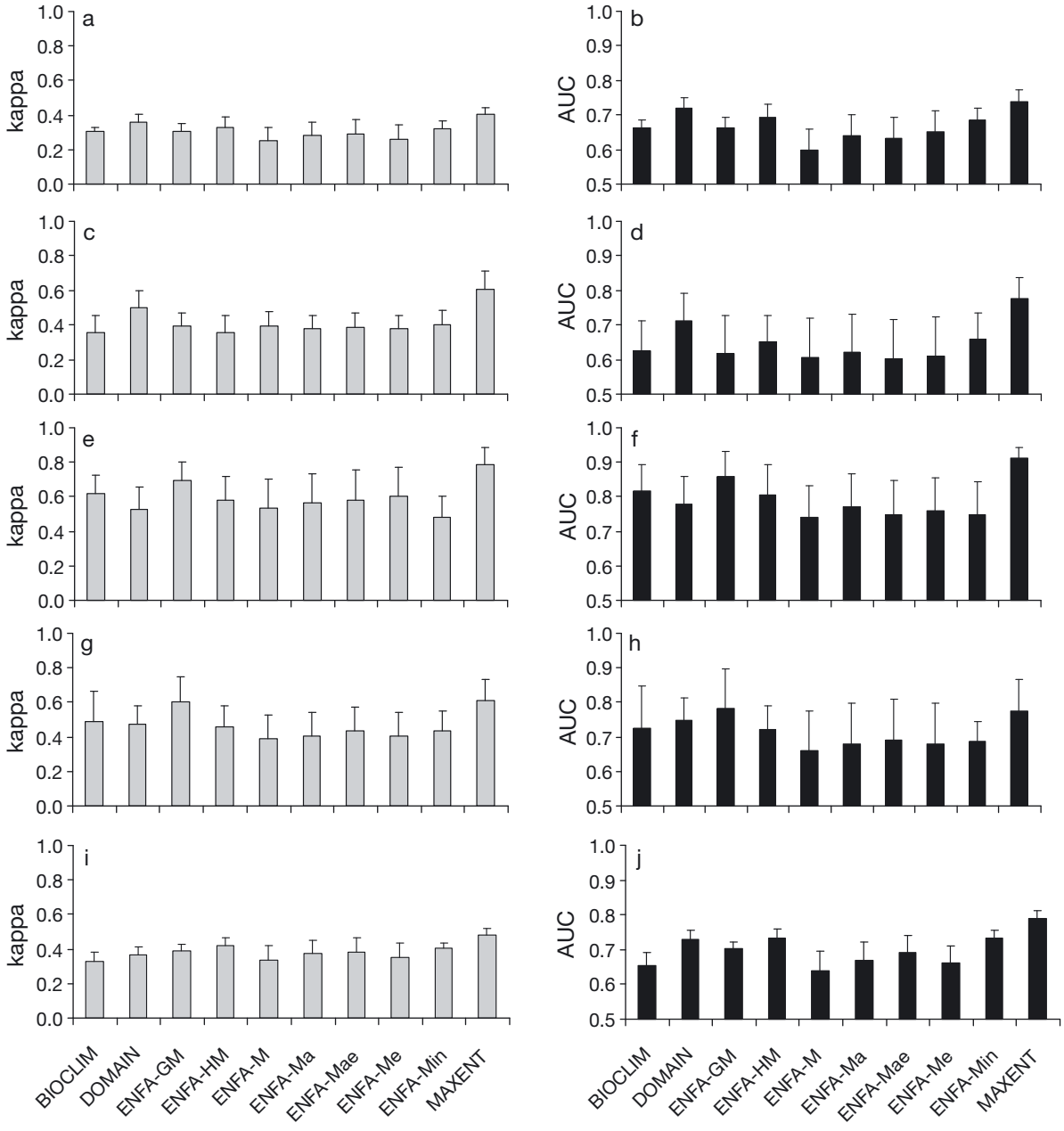

Fig. 3. Model performance for (a,b) *Caesioperca* spp.: (a) kappa, pseudo-$F_{9,81}$ = 8.95, p < 0.001, (b) AUC (area under curve), pseudo-$F_{9,81}$ = 13.93, p < 0.001; (c,d) *Cheilodactylus nigripes*: (c) kappa, pseudo-$F_{9,81}$ = 7.64, p < 0.001, (d) AUC, pseudo-$F_{9,81}$ = 4.57, p < 0.001; (e,f) *Notolabrus tetricus*: (e) kappa, pseudo-$F_{9,81}$ = 6.85, p < 0.001, (f) AUC, pseudo-$F_{9,81}$ = 6.07, p < 0.001; (g,h) *Pempheris multiradiata*: (g) kappa, pseudo-$F_{9,81}$ = 6.63, p < 0.001, (h) AUC, pseudo-$F_{9,81}$ = 3.41, p < 0.001; and (i,j) *Pseudolabrus psittaculus*: (i) kappa, pseudo-$F_{9,81}$ = 9.74, p < 0.001, (j) AUC, pseudo-$F_{9,81}$ = 17.10, p < 0.001 produced with models BIOCLIM, DOMAIN, ENFA-M, -Ma, -Mae, -Me, -GM, -HM, -Min and MAXENT (see 'Materials and methods: modelling approaches'). Error bars: SD

tance to nearest reef, HSI-b and maximum curvature (Table 5). Examination of the response curves for these variables showed that areas delineated by depths ~40 m, close to reef, with low rugosity, intermediate maximum curvature and high HSI-b values were the most suitable habitats for this species (Fig. 5).

The shallow (<20 m), highly complex reef structures in the eastern region of the study site were found to have the highest habitat suitability for *Notolabrus tetricus* (Fig. 4f). Some fragmented, highly suitable habitat was also predicted throughout the deeper (>50 m) regions in the north-west of the study site. Among

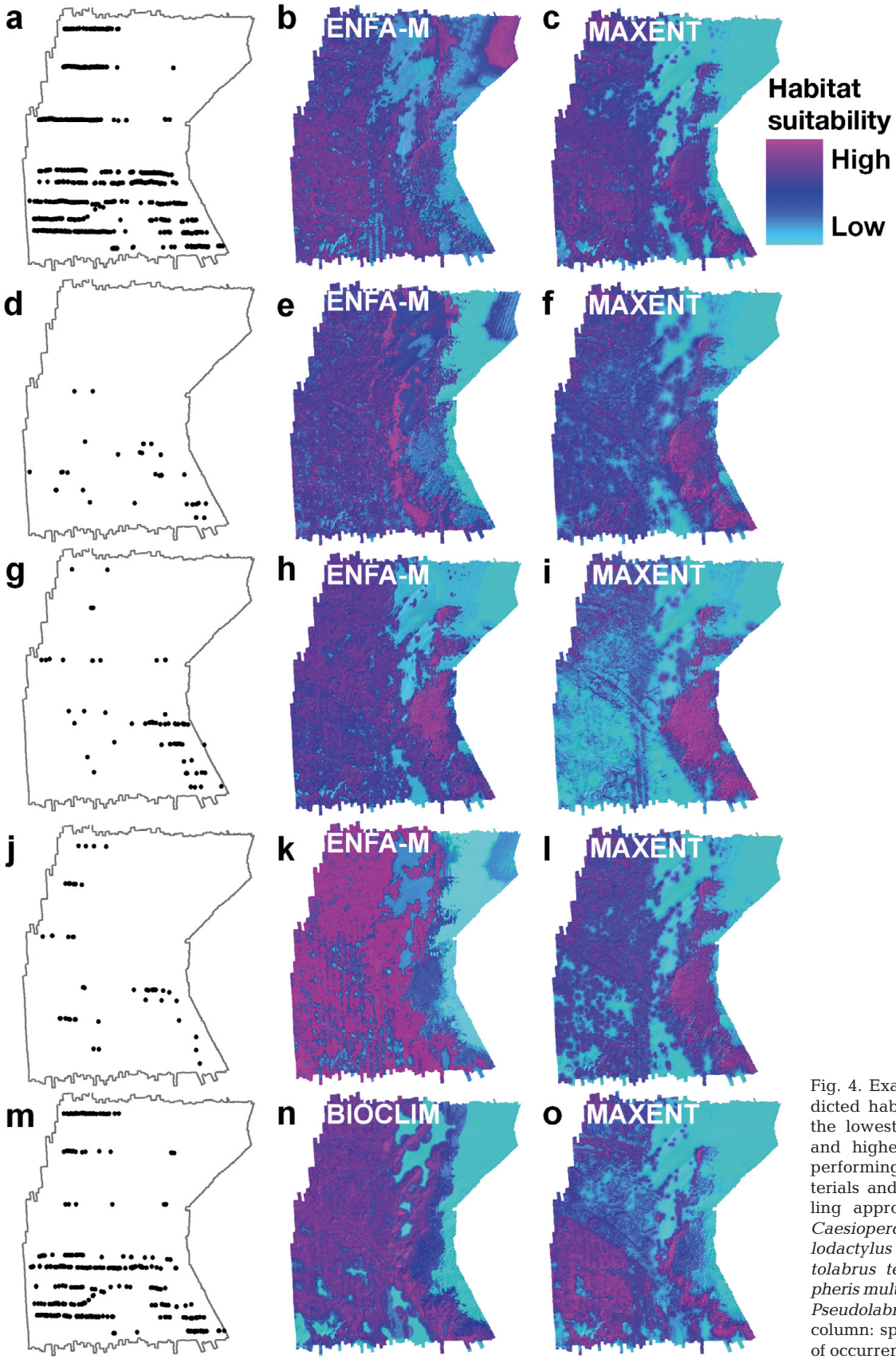

Fig. 4. Examples of the predicted habitat suitability for the lowest (middle column) and highest (right column) performing models (see 'Materials and methods: modelling approaches') for (a–c) *Caesioperca* spp., (d–f) *Cheilodactylus nigripes*, (g–i) *Notolabrus tetricus*, (j–l) *Pempheris multiradiata* and (m–o) *Pseudolabrus psittaculus*. Left column: spatial arrangement of occurrences for each taxon

Table 4. Summary of orthogonal contrasts used in determining influence of species range and environmental niches on model performance. AUC: area under the curve; AOO: area of occupancy. Significant contrasts shown **in bold**

Contrast description	Contrast	*t*-statistic (45 df)	Significance (1-tailed)
Marginality influence on mean kappa	1	**6.298**	**<0.001**
	2	**4.320**	**<0.001**
	3	1.292	0.105
	4	**5.818**	**<0.001**
Marginality influence on mean AUC	1	**2.844**	**0.004**
	2	1.174	0.123
	3	0.455	0.326
	4	**6.630**	**<0.001**
Tolerance influence on mean kappa	1	1.041	0.152
	2	–1.579	0.061
	3	**–9.209**	**<0.001**
	4	**–2.333**	**0.012**
Tolerance influence on mean AUC	1	**3.944**	**<0.001**
	2	0.322	0.375
	3	**–5.969**	**<0.001**
	4	–1.530	0.067
AOO influence on mean kappa	1	1.041	0.152
	2	–1.579	0.061
	3	**–9.209**	**<0.001**
	4	**–2.333**	**0.012**
AOO influence on mean AUC	1	**3.994**	**<0.001**
	2	0.322	0.375
	3	**–5.969**	**<0.001**
	4	–1.530	0.065

Table 5. Relative variable importance as measured by regularised gain for each fish taxon. BPI: benthic position index; HSI-b: Hue-saturation-intensity (blue). Variables contributing most to predictions are shown **in bold**

Taxon	Variable	Regularised gain
Caesioperca spp.		
	Bathymetry	**0.16**
	BPI	0.06
	Eastness	0.05
	Euclidean distance to nearest reef	**0.13**
	HSI-b	**0.13**
	Max. curvature	0.08
	Northness	0.03
	Rugosity	**0.11**
Cheilodactylus nigripes		
	Bathymetry	0.06
	BPI	0.09
	Eastness	**0.10**
	Euclidean distance to nearest reef	**0.27**
	HSI-b	**0.13**
	Max. curvature	**0.10**
	Northness	0.05
	Rugosity	**0.39**
Notolabrus tetricus		
	Bathymetry	**0.39**
	BPI	**0.15**
	Eastness	0.03
	Euclidean distance to nearest reef	**1.00**
	HSI-b	**0.29**
	Max. curvature	**0.11**
	Northness	0.03
	Rugosity	**0.89**
Pempheris multiradiata		
	Bathymetry	0.05
	BPI	0.02
	Eastness	**0.10**
	Euclidean distance to nearest reef	**0.35**
	HSI-b	**0.14**
	Max. curvature	**0.19**
	Northness	**0.13**
	Rugosity	**0.17**
Pseudolabrus psittaculus		
	Bathymetry	**0.24**
	BPI	**0.11**
	Eastness	0.05
	Euclidean distance to nearest reef	**0.14**
	HSI-b	**0.17**
	Max. curvature	**0.15**
	Northness	0.03
	Rugosity	**0.17**

seafloor variables, Euclidean distance to nearest reef, rugosity, bathymetry, HSI-b, BPI and maximum curvature were selected as important variables for defining habitat suitability for this species (Table 5). Examination of the response curves for these variables highlighted regions that were delineated by close proximity to reef, high rugosity, high HSI-b, either highly positive or negative BPI (i.e. peaks or troughs), intermediate maximum curvature, and ~20 m water depth as the most suitable habitat for *N. tetricus* (Fig. 5).

The predicted distribution of suitable habitat for *Pempheris multiradiata* showed similar patterns to that of *Notolabrus tetricus*, albeit more widespread in the deeper regions of the study area (Fig. 4h). Several seafloor variables defined the predicted distribution of suitable habitat for *P. multiradiata*: Euclidean distance to nearest reef, maximum curvature, rugosity, HSI-b, northness and eastness (Table 5). Response curves of these variables indicated that, in general, south-east facing (i.e. northness values ca. –0.8 and eastness values ca. 0.8) regions close to reef, with low rugosity, moderate maximum curvature and high HSI-b values were the best indicators for suitable habitat for this species (Fig. 5).

Pseudolabrus psittaculus was predicted by MAXENT to have similar patterns to *Cheilodactylus nigripes*, with highly suitable habitat being confined to the deeper regions of the study area (Fig. 4j). Analysis of jackknife results showed that bathymetry, rugosity, HSI-b, maximum curvature, Euclidean distance to nearest reef and BPI were the primary determinants in characterising the habitat suitability for this species (Table 5). Response curves of these seafloor variables indicated that the highest habitat suitability for *P. psittaculus* was observed with the increase in depth (Fig. 5a). However, the curves showed this trend up to certain point (~31 m depth) and beyond that the habi-

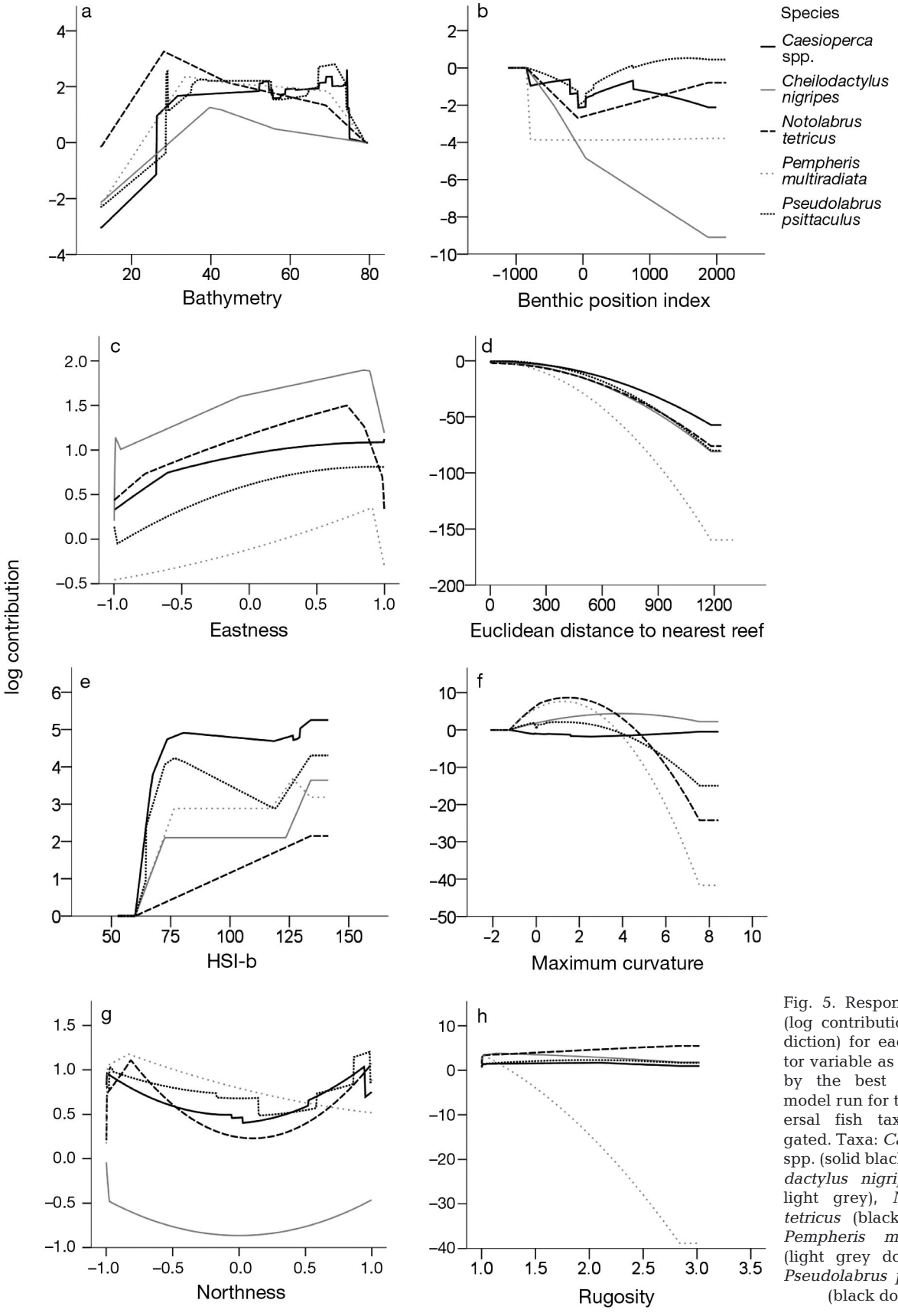

Fig. 5. Response curves (log contribution to prediction) for each predictor variable as generated by the best MAXENT model run for the 5 demersal fish taxa investgated. Taxa: *Caesioperca* spp. (solid black), *Cheilodactylus nigripes* (solid light grey), *Notolabrus tetricus* (black dashed), *Pempheris multiradiata* (light grey dotted) and *Pseudolabrus psittaculus* (black dotted)

tat suitability plateaued until 73 m after which there was a sharp decline. The response curve for rugosity (Fig. 5h) showed a similar trend to that of *Caesioperca* spp. with only a slight increase in habitat suitability of *P. psittaculus* around rugosity values of 2. Lower values of HSI-b were associated with most suitable habitat for *P. psittaculus* (Fig. 5e). Increased maximum curvature decreased habitat suitability (Fig. 5f). Euclidean distance to nearest reef showed the same trend as all other taxa, with diminishing suitability increasing with distance from reef (Fig. 5d). Albeit more pronounced than for *Notolabrus tetricus,* BPI showed a decrease in suitable habitat around BPI values of 0 for *P. psittaculus* (Fig 5b).

DISCUSSION

Predictive performance

This study compared 10 presence-only algorithms to predict the habitat suitability of 5 demersal fish taxa in Discovery Bay, Australia. To the best of our knowledge, this is the first time that towed-video-derived fish occurrence and detailed spatially-explicit seafloor datasets have been used to compare presence-only modelling approaches. Overall, we found that the range in kappa and AUC values from the models were comparable with other marine and terrestrial presence-only modelling studies (e.g. Elith et al. 2006, Tsoar et al. 2007, Tittensor et al. 2009). Statistically significant differences in values of kappa and AUC were observed between the modelling algorithms. Generally, MAXENT significantly outperformed the other algorithms, similar to findings from previous studies. Tittensor et al. (2009) used MAXENT and ENFA-GM to predict global habitat suitability of stony corals on seamounts and found that MAXENT consistently outperformed ENFA-GM. Similarly, Elith et al. (2006), who compared 16 modelling approaches using a variety of terrestrial flora and fauna from 6 different regions of the world, found that MAXENT significantly outperformed BIOCLIM and DOMAIN. We also found that there was no significant difference between the more recently developed ENFA-Ma, ENFA-Mae, ENFA-Me approaches and the original ENFA-M algorithm. This is in contrast to findings by Braunisch et al. (2008), who found that the more recent ENFA algorithms provided better models of habitat suitability for large forest grouse *Tetrao urogallus* than the original ENFA-M algorithm.

There are numerous reasons for the observed differences in predictive model performances in this study. Differences could potentially be due to the modelling algorithms ability to fit the complex species–environment relationships. Pearce & Ferrier (2000) suggested that algorithms used to fit species distribution models can be ranked in accordance to their 'function complexity'. The models examined in our study differ considerably from each other in their complexity. BIOCLIM is conceptually the simplest model tested, assuming a rectilinear environmental envelope around occurrence data in environmental space. Consequently, only the outer records along each environmental variable are used to define the boundaries of the ecological niche and the model cannot deal with correlations or interactions between the environmental variables (Hijmans et al. 2005). DOMAIN differs from BIOCLIM in its ability to cope with discontinuity of the species occurrence datasets in environmental space (Hijmans et al. 2005). Its main restraint is that, for each potential site, only a single record (the nearest neighbour in environmental space) is used to determine suitable habitat. The original ENFA-related algorithms (GM, HM, M and Min) take into account the distribution of all the occurrence records in environmental space and create elliptic envelopes that are consistent with the assumption of unimodal responses to environmental gradients. The more recent ENFA algorithms (Ma, Me, Mae) are slightly more complex than the original ENFA algorithms, taking into account marginality and specialisation (see Braunisch et al. 2008). Because of MAXENT's ability to iteratively evaluate and improve the rules used for generating predictions, it is the most complex algorithm among the methods examined in this study. Our results indicate that the 10 modelling algorithms can be generalised into 3 groups based on their kappa and AUC scores: (1) MAXENT consistently performed best; (2) ENFA-GM, ENFA-HM and DOMAIN had intermediate levels of performance; and (3) BIOCLIM, ENFA-M, ENFA-Min and the more recent ENFA algorithms performed lowest. Consequently, with the exception of the relatively poor performance obtained for the more recent ENFA algorithms, our results are consistent with the hypothesis that increasing model complexity may provide a better fit for complex species–environment relationships, resulting in better model performance.

Influence of species' range and environmental niches on modelling performance

Niche position (expressed by marginality) was the only descriptor to be significantly associated with model performance (i.e. kappa). We obtained higher performing models based on occurrence data from fish taxa with niche positions that were greatly different from the mean conditions of the study area. However, no discernible association was achieved from

the other environmental niche (i.e. tolerance) and species range (i.e. AOO) descriptors examined. In relation to marginality, this association with model performance (kappa) coincides with observations made by Hernandez et al. (2006), who found that predictive performance of organisms with clearly definable environmental niches (i.e. high marginality) can be modelled with higher performance than those of more generalist species (i.e. low marginality). Our lack of association between model performance and the remaining descriptors contrasts with previous terrestrial studies. Tsoar et al. (2007) found that tolerance had a negative effect on predictive performance as measured by kappa. A possible explanation for these discrepancies between our findings and previous studies could relate to the spatial scale at which the variables are generated and fitted. Brotons et al. (2004) proposed that species inhabiting a wide range of habitats in a certain area might not be limited by any of the measured variables at the scale at which the models are fitted and may therefore perform lower. Because this study was primarily concerned with assessing model performance between algorithms, the same variables were used across all taxa and were generated at the finest possible spatial scale (3 × 3 cell analysis window). We acknowledge that there is no universally correct spatial scale at which to describe species–habitat relationships (Wiens 1989) and that better models may be achieved if variables were generated at multiple spatial scales.

Most important variables for characterisation of suitable habitat for demersal fishes

The integration of occurrence data from towed-video and multibeam sonar-derived seafloor variables in a presence-only modelling framework were useful for explaining the characteristics that define suitable habitat for the 5 demersal fish taxa investigated. We found that MAXENT models consistently delineated areas of suitable habitat to be regions of seafloor that relate to the known ecology of these focal taxa. For example, adult *Caesioperca* spp. are known to school on top of deeper (>30 m) rocky outcrops, feeding on plankton, while juveniles remain close to the rocky substrata (Edgar 2000, Williams & Bax 2001). Its distribution of suitable habitat was defined with seafloor variables that express these deeper reef drop-offs. Similarly, *Cheilodactylus nigripes* is a common species that inhabits reefs, over a range of depths, feeding on benthic invertebrates (Gomon et al. 2008). Highly suitable habitat was identified as the less complex intermediate to deep reef systems, which are known to be densely covered in a variety of sponges, ascidians and bryozoans (Ierodiaconou et al. 2007a). *Pempheris multiradiata* are generally observed near over-hangs and caves near the edges of reefs (Edgar 2000). Consistent with this, the MAXENT model highlighted areas of highly suitable habitat as the protected south-east facing reefs (the predominant swell direction is from the south-west) with low bathymetric complexity.

The MAXENT models highlighted an important difference in the ecological requirements of 2 wrasse species (Labridae). While both *Notolabrus tetricus* and *Pseudolabrus psittaculus* are benthic carnivores that consume a variety of mobile and sessile invertebrates (Gomon et al. 2008), differences in seafloor variables that influence their habitat suitability were observed. *N. tetricus* juveniles and small females are commonly observed in kelp- and seagrass-dominated shallow waters (<20 m), while larger fish inhabit deeper invertebrate-dominated rocky reefs (Edgar 2000). Males of this species are fiercely territorial and maintain a harem of females over home ranges of 400 to 775 m^2 (Barrett 1995). By contrast, *P. psittaculus* were commonly observed in pairs, or small schools, in and around sessile invertebrate- and thallose-red-algal-dominated reef systems that are >30 m. Examining the predicted habitat suitability and response curves of these 2 species indicated that highly suitable habitat for *N. tetricus* was predicted as present in shallow areas of reef with high bathymetric complexity (i.e. high rugosity and either positive or negative BPI); whilst *P. psittaculus* was predicted to be mainly confined to the less complex deeper reefs.

Many studies have identified the influence of seafloor characteristics on the distributions of suitable habitat for demersal fishes because they are either direct or indirect proxies that represent important physiological or ecological limitations, including the availability of territory, food, shelter or the existence of predation or competition (Choat & Ayling 1987, Friedlander & Parrish 1998, Priede & Merrett 1998, García-Charton et al. 2004, Moore et al. 2009, Chatfield et al. 2010). Our study, however, highlights the importance of having detailed spatially explicit (i.e. full-coverage) seafloor data rather than the point-located descriptors relied on by earlier studies (Friedlander & Parrish 1998, Babcock et al. 1999, Westera et al. 2003, Willis & Anderson 2003). These spatially continuous measures of seafloor data reflect subtle, but important, differences in habitat suitability that provide end users (e.g. management agencies or research scientists) with accurate and detailed spatially explicit information about demersal fishes. Consequently, the spatially explicit model predictions generated in this study should be viewed as a framework upon which targeted empirical research can be based.

CONCLUSIONS

The comparison of 10 algorithms provided a comprehensive evaluation of which presence-only techniques are most suited to modelling suitable habitat of demersal fishes. Generally, MAXENT produced the best performing models for the taxa and study area examined. We also found that fish with clearly definable environmental niches can be modelled with higher performance than those of more generalist species. The continued use of these presence-only models, particularly MAXENT, is encouraged. Further comparisons, however, with other approaches, including presence/absence methods (using pseudo-absence data), are also encouraged. MAXENT models clearly demonstrated the value of remotely-sensed occurrence and detailed spatially-explicit seafloor datasets in determining the importance of variables that influence suitable habitat of demersal fishes.

While the models are still only describing associations and not necessarily demonstrating causal relationships, it is their ability to develop realistic response curves and to provide spatially-explicit predictions of suitable habitat that makes these approaches a useful management tool. There are, however, a number of issues and limitations that need to be taken into account if model predictions are to be incorporated into management decisions. First, compared to presence/absence approaches, presence-only models inevitably tend to over-predict suitable habitat (Zaniewski et al. 2002). This is because there is no absence data to constrain the model predictions of suitable habitat. Overestimating suitable habitat, however, might be more preferable to underestimating its existence, particularly when considering commercially- or ecologically important species that are likely to be the focus of management interventions (Fielding & Bell 1997). Indeed, it is commonly perceived that presence-only predictions reflect potential habitat suitability, whereas presence/absence methods reflect the present habitat suitability of the species (Soberón & Peterson 2007, Chefaoui & Lobo 2008, Jiménez-Valverde et al. 2008). Second, our models are based on a one-off survey. Many fish species exhibit temporal migration patterns between habitats and regions (Malcolm et al. 2007). Ideally, additional surveys and models are needed to quantify temporal changes in habitat preferences. Third, although it is notable that models had strong predictive performances, even when restricted to seafloor variables, we suspect that model performance could be improved with the inclusion of additional explanatory variables such as wave exposure and prey availability (if and when they become available). Finally, habitat ontogenies among fishes are well documented (Green 1996, Gratwicke et al. 2006). While not attainable from the current camera system, the use of size class information (obtained from stereo video methods; e.g. Harvey et al. 2002) in models may also provide more accurate predictions of suitable habitat and potentially highlight differences in size class/life stage habitat dependency. Despite these issues, the model predictions presented in this study provide an accurate baseline that may assist researchers in addressing more targeted biological questions and helping managers and conservation practitioners to ensure that marine resources are managed in a sustainable manner.

Acknowledgements. This work was supported by National Heritage Trust/Caring for our Country and Parks Victoria as part of the Victorian Marine Habitat Mapping Project. Thanks to the crew of the Australian Maritime College vessel 'Bluefin' and Fugro Survey for the collection of multibeam sonar data. Thanks to the crew aboard Deakin University vessel 'Courageous II' for undertaking the video survey. Spatial analyses were undertaken at Deakin University, Warrnambool, Victoria, GIS Laboratory facility. Thanks also the 6 anonymous reviewers and the editor whose comments greatly improved this paper.

LITERATURE CITED

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32–46

Anderson TJ, Yoklavich MM (2007) Multiscale habitat associations of deepwater demersal fishes off central California. Fish Bull 105:168–179

Araújo MB, Williams PH (2000) Selecting areas for species persistence using occurrence data. Biol Conserv 96: 331–345

Assis J, Narváez K, Haroun R (2007) Underwater towed video: a useful tool to rapidly assess elasmobranch populations in large marine protected areas. J Coast Conserv 11:153–187

Babcock RC, Kelly S, Shears NT, Walker JW, Willis TJ (1999) Changes in community structure in temperate marine reserves. Mar Ecol Prog Ser 189:125–134

Barrett NS (1995) Short- and long-term movement patterns of six temperate reef fishes (Families Labridae and Monacanthidae). Mar Freshw Res 46:853–860

Boutakoff N (1963) The geology and geomorphology of the Portland area. Department of Mines, Melbourne

Braunisch V, Bollmann K, Graf RF, Hirzel AH (2008) Living on the edge: modelling habitat suitability for species at the edge of their fundamental niche. Ecol Model 214:153–167

Brotons L, Thuiller W, Araújo MB, Hirzel AH (2004) Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27: 437–448

Busby JR (1986) A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook. Oerst.) in southeastern Australia. Austral Ecol 11:1–7

Cañadas A, Sagarminaga R, Stephanis RD, Urquiola E, Hammond PS (2005) Habitat preference modelling as a conservation tool: proposals for marine protected areas for cetaceans in southern Spanish waters. Aquat Conserv Mar Freshw Ecosyst 15:495–521

Carpenter G, Gillison AN, Winter J (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants, animals. Biodivers Conserv 2:667–680

Chatfield BS, Van Niel KP, Kendrick GA, Harvey ES (2010) Combining environmental gradients to explain and predict the structure of demersal fish distributions. J Biogeogr 37:593–605

Chefaoui RM, Lobo JM (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. Ecol Model 210:478–486

Choat JH, Ayling AM (1987) The relationship between habitat structure and fish faunas on New Zealand reefs. J Exp Mar Biol Ecol 110:257–284

Clarke K, Gorley R (2006) PRIMER v6 user manual/tutorial. PRIMER-E, Plymouth

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46

Compton TJ, Leathwick JR, Inglis GJ (2010) Thermogeography predicts the potential global range of the invasive European green crab (*Carcinus maenas*). Divers Distrib 16:243–255

Daily M (1983) Hue-saturation-intensity split-spectrum processing of Seasat radar imagery. Photogramm Eng Remote Sensing 49:349–355

Edgar G (2000) Australian marine life: the plants and animals of temperate waters. Reed Books, Sydney

Elith J, Burgman MA (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: Scott JM, Heglund PJ, Morrison ML, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences: issues of scale and accuracy. Island Press, Covelo, CA

Elith J, Graham CH, Anderson RP, Dudik M and others (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

Engler R, Guisan A, Rechsteiner L (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J Appl Ecol 41:263–274

Farber O, Kadmon R (2003) Assessment of alternative approaches for bioclimatic modelling with special emphasis on the Mahalanobis distance. Ecol Model 160: 115–130

Ferrier S, Drielsma M, Manion G, Watson G (2002a) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. Biodivers Conserv 11:2309–2338

Ferrier S, Watson G, Pearce J, Drielsma M (2002b) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodivers Conserv 11:2275–2307

Field A (2000) Discovering statistics using SPSS for Windows. Sage Publications, London

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Franklin J (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Prog Phys Geogr 19:474–499

Friedlander AM, Parrish JD (1998) Habitat characteristics affecting fish assemblages on a Hawaiian coral reef. J Exp Mar Biol Ecol 224:1–30

Gallant JC, Wilson JP (1996) TAPES-G: a grid-based terrain analysis program for the environmental sciences. Comput Geosci 22:713–722

Galparsoro I, Borja Á, Bald J, Liria P, Chust G (2009) Predicting suitable habitat for the European lobster (*Homarus gammarus*), on the Basque continental shelf (Bay of Biscay), using Ecological-Niche Factor Analysis. Ecol Model 220:556–567

García-Charton JA, Pérez-Ruzafa Á, Sánchez-Jerez P, Bayle-Sempere JT, Reñones O, Moreno D (2004) Multi-scale spatial heterogeneity, habitat structure, and the effect of marine reserves on western Mediterranean rocky reef fish assemblages. Mar Biol 144:161–182

Gaston KJ (1996) Species-range-size distributions: patterns, mechanisms and implications. Trends Ecol Evol 11: 197–201

Gomon MF, Bray D, Kuiter RH (eds) (2008) The fishes of Australia's South Coast. New Holland, Sydney

Gratwicke B, Petrovic C, Speight MR (2006) Fish distribution and ontogenetic habitat preferences in non-estuarine lagoons and adjacent reefs. Environ Biol Fishes 76: 191–210

Green AL (1996) Spatial, temporal and ontogenetic patterns of habitat use by coral reef fishes (Family Labridae). Mar Ecol Prog Ser 133:1–11

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. Conserv Biol 20:501–511

Harvey E, Fletcher D, Shortis M (2002) Estimation of reef fish length by divers and by stereo-video: a first comparison of the accuracy and precision in the field on living fish under operational conditions. Fish Res 57:255–265

Hernandez PA, Graham C, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modelling methods. Ecography 29:773–785

Hijmans RJ, Graham CH (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. Glob Change Biol 12:2272–2281

Hijmans RJ, Guarino L, Jarvis A, O'Brien R and others (2005) DIVA-GIS Version 5.2 user manual. LizardTech, available at: www.diva-gis.org/docs/DIVA-GIS5_manual.pdf

Hirzel AH, Helfer W, Metral F (2001) Assessing habitat-suitability models with a virtual species. Ecol Model 145: 111–121

Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? Ecology 83:2027–2036

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Model 199:142–152

Hirzel AH, Hausser J, Perrin N (2007) Biomapper 4.0 user manual. University of Lausanne, Lausanne

Hutchinson GE (1957) Concluding remarks. Cold Spring Harbour Symp Quant Biol 22:415–427

Iampietro PJ, Kvitek RG, Morris E (2005) Recent advances in automated genus-specific marine habitat mapping enabled by high-resolution multibeam bathymetry. Mar Technol Soc J 39:83–93

Ierodiaconou D, Rattray A, Laurenson L, Monk J, Lind P (2007a) Victorian marine habitat mapping project. Deakin University, Warrnambool

Ierodiaconou D, Laurenson L, Burq S, Reston M (2007b) Marine benthic habitat mapping using multibeam data, georeferenced video and image classification techniques in Victoria, Australia. J Spatial Sci 52:93–104

Ierodiaconou D, Monk J, Rattray A, Laurenson L, Vercase VL (2010) Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. Cont Shelf Res (in press) doi:10.1016/j.csr.2010.01.012

Jiménez-Valverde A, Lobo JM, Hortal J (2008) Not as good as they seem: the importance of concepts in species distribution modelling. Divers Distrib 14:885–890

Langlois TJ, Harvey ES, Fitzpatrick B, Meeuwig JJ, Shedrawi G, Watson DL (2010) Cost-efficient sampling of fish assemblages: comparison of baited video stations and diver video transects. Aquat Biol 9:155–168

Li J, Hilbert D (2008) LIVES: a new habitat modelling technique for predicting the distribution of species' occurrences using presence-only data based on limiting factor theory. Biodivers Conserv 17:3079–3095

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17:145–151

Loiselle BA, Howell CA, Graham CH, Goerck JM, Brooks T, Smith KG, Williams PH (2003) Avoiding pitfalls of using species distribution models in conservation planning. Conserv Biol 17:1591–1600

Lundblad ER, Wright DJ, Miller J, Larkin EM and others (2006) A Benthic terrain classification scheme for American Samoa. Mar Geod 29:89–111

MacArthur RH (1957) On the relative abundance of bird species. Proc Natl Acad Sci USA 43:293–295

MacLeod CD, Mandleberg L, Schweder C, Bannon SM, Pierce GJ (2008) A comparison of approaches for modelling the occurrence of marine animals. Hydrobiologia 612: 21–32

Malcolm HA, Gladstone W, Lindfield S, Wraith J, Lynch TP (2007) Spatial and temporal variation in reef fish assemblages of marine parks in New South Wales, Australia—baited video observations. Mar Ecol Prog Ser 350:277–290

Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. J Appl Ecol 38:921–931

Monk J, Ierodiaconou D, Bellgrove A, Harvey E, Laurenson L (2010) Remotely sensed hydroacoustics and observation data for predicting fish habitat suitability. Cont Shelf Res (in press) doi:10.1016/j.csr.2010.02.012

Moore CH, Harvey ES, Van Niel KP (2009) Spatial prediction of demersal fish distributions: enhancing our understanding of species–environment relationships. ICES J Mar Sci 66:2068–2075

Morrison M, Carbines G (2006) Estimating the abundance and size structure of an estuarine population of the sparid *Pagrus auratus*, using a towed camera during nocturnal periods of inactivity, and comparisons with conventional sampling techniques. Fish Res 82:150–161

Mueller R, Brown R, Hop H, Moulton L (2006) Video and acoustic camera techniques for studying fish under ice: a review and comparison. Rev Fish Biol Fish 16:213–226

Murphy HM, Jenkins GP (2010) Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. Mar Freshw Res 61:236–252

Nasby-Lucas NM, Embley BW, Hixon MA, Merle SG, Tissot BN, Wright DJ (2002) Integration of submersible transect data and high-resolution multibeam sonar imagery for a habitat-based groundfish assessment of Heceta Bank, Oregon. Fish Bull 100:739–751

Nix H (1986) A biogeographic analysis of Australian elapid snakes. In: Longmore R (ed) Atlas of elapid snakes of Australia. Australian Government Publishing Service, Canberra

Olden JD, Jackson DA (2002) A comparison of statistical approaches for modelling fish species distributions. Freshw Biol 47:1976–1995

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133:225–245

Pearce J, Ferrier S (2001) The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biol Conserv 98:33–43

Pearce J, Ferrier S, Scotts D (2001) An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. J Environ Manage 62:171–184

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modelling of species geographic distributions. Ecol Model 190:231–259

Pittman SJ, Caldow C, Hile SD, Monaco ME (2007) Using seascape types to explain the spatial patterns of fish in the mangroves of SW Puerto Rico. Mar Ecol Prog Ser 348: 273–284

Pittman SJ, Costa BM, Battista TA (2009) Using lidar bathymetry and boosted regression trees to predict the diversity and abundance of fish and corals. J Coast Res 53:27–38

Priede IG, Merrett NR (1998) The relationship between numbers of fish attracted to baited cameras and population density: studies on demersal grenadiers *Coryphaenoides* (*Nematonurus*) *armatus* in the abyssal NE Atlantic Ocean. Fish Res 36:133–137

Rattray A, Ierodiaconou D, Laurenson L, Burq S, Reston M (2009) Hydro-acoustic remote sensing of benthic biological communities on the shallow South East Australian continental shelf. Estuar Coast Shelf Sci 84:237–245

Roberts DW (1986) Ordination on the basis of fuzzy set theory. Vegetatio 66:123–131

Robertson MP, Caithness N, Villet MH (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. Divers Distrib 7:15–27

Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. J Biogeogr 31:1555–1568

Sérgio C, Figueira R, Draper D, Menezes R, Sousa AJ (2007) Modelling bryophyte distribution based on ecological information for extent of occurrence assessment. Biol Conserv 135:341–351

Soberón J, Peterson T (2004) Biodiversity informatics: managing and applying primary biodiversity data. Philos Trans R Soc Lond B 359:689–698

Spencer ML, Stoner AW, Ryer CH, Munk JE (2005) A towed camera sled for estimating abundance of juvenile flatfishes and habitat characteristics: comparison with beam trawls and divers. Estuar Coast Shelf Sci 64:497–503

Stockwell DRB, Noble IR (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data-analysis. Math Comput Simul 33:385–390

Thuiller W (2003) BIOMOD: optimizing predictions of species distributions and projecting potential future shifts under global change. Glob Change Biol 9:1353–1362

Tittensor DP, Baco AR, Brewin PE, Clark MR and others (2009) Predicting global habitat suitability for stony corals on seamounts. J Biogeogr 36:1111–1128

Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R (2007) A comparative evaluation of presence-only methods for modelling species distribution. Divers Distrib 13:397–405

Watson DL, Harvey ES, Anderson MJ, Kendrick GA (2005) A comparison of temperate reef fish assemblages recorded by three underwater stereo-video techniques. Mar Biol 148:415–425

Wedding LM, Friedlander AM, McGranaghan M, Yost RS, Monaco ME (2008) Using bathymetric Lidar to define nearshore benthic habitat complexity: implications for management of reef fish assemblages in Hawaii. Remote Sens Environ 112:4159–4165

Weiss AD (2001) Topographic positions and landforms analysis. ESRI International User Conference, San Diego, CA, 9–11 July

Westera M, Lavery P, Hyndes G (2003) Differences in recreationally targeted fishes between protected and fished areas of a coral reef marine park. J Exp Mar Biol Ecol 294:145–168

Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3: 385–397

Wiley EO, McNyset KM, Townsend Peterson A, Richard Robins C, Stewart AM (2003) Niche modelling and geographic range predictions in the marine environment using a machine-learning algorithm. J Oceanogr 16: 120–127

Williams A, Bax NJ (2001) Delineating fish-habitat associations for spatially based management: an example from the south-eastern Australian continental shelf. Mar Freshw Res 52:513–536

Williams GJ, Aeby GS, Cowie ROM, Davy SK (2010) Predictive modelling of coral disease distribution within a reef system. PLoS ONE 5:e9264

Willis TJ, Anderson MJ (2003) Structure of cryptic reef fish assemblages: relationships with habitat characteristics and predator density. Mar Ecol Prog Ser 257:209–221

Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biol Conserv 122:99–112

Wilson MFJ, O'Connell B, Brown C, Guinan JC, Grehan AJ (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Mar Geod 30:3–35

Ysebaert T, Meire P, Herman PMJ, Verbeek H (2002) Macrobenthic species response surfaces along estuarine gradients: prediction by logistic regression. Mar Ecol Prog Ser 225:79–95

Zaniewski AE, Lehmann A, Overton JMC (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol Model 157: 261–280

Editorial responsibility: Jake Rice, Ottawa, Ontario, Canada

Submitted: July 5, 2010; Accepted: October 3, 2010
Proofs received from author(s): December 6, 2010