



UNIVERSIDADE DOS AÇORES
DEPARTAMENTO DE OCEANOGRAFIA E PESCAS

DISSERTAÇÃO DE MESTRADO EM ESTUDOS INTEGRADOS DOS OCEANOS

HABITAT PREDICTIVE MODELLING OF DEMERSAL FISH SPECIES IN THE AZORES

Hugo Alexandre Esteves Parra

Orientador: Telmo Morato

Co-orientadores: Gui Menezes e Fernando Tempera

Horta, Novembro de 2012

ACKNOWLEDGMENTS

I would like to send my gratitude to everybody who supported me in accomplishing this task, especially to:

My supervisor Telmo Morato for his guidance, encouragement, valuable support on my thesis writing and for giving me the opportunity to meet new places and interesting people in a foreign conference. It was definitely a very enriching experience that I'll never forget.

My co-supervisors Gui Menezes and Fernando Tempera for providing the needed information and suggestions to develop this work.

DOP investigators Xana Guedes, Patricia Amorim and Ricardo Medeiros for the precious technical help.

The Captain and crew of the research vessel "N/I Arquipélago" and to all that collaborate in demersal surveys.

Paula Lourinho for providing me assistance in one way or another, especially when my laptop screen broke early morning outside the department, the day before the dead line to submit this work. This is something I'll laugh one day.

Last but not the least, my family and friends which helped me reach where I am today. A special thanks to Fernanda and Lucas for their support when I needed the most in this last year. To all, mentioned or not, my most sincere thank you.

ABSTRACT

Species distribution modelling of the marine environment has been extensively used to assess species–environment relationships to predict fish spatial distributions accurately. In this study we explored the application of two distinct modelling techniques, maximum entropy model (MaxEnt) and generalized linear models (GLMs) for predicting the potential distribution in the Azores economic exclusive zone (EEZ) of four economically important demersal fish species: blackbelly rosefish, *Helicolenus dactylopterus dactylopterus*, forkbeard, *Phycis phycis*, wreckfish, *Polyprion americanus* and offshore rockfish, *Pontinus kuhlii*. Models were constructed based on 13 years of fish presence/absence data derived from bottom longline surveys performed in the study area combined with high resolution (300 m) topographic and biogeochemical habitat seafloor variables. The most important predictors were depth and slope followed by sediment type, oxygen saturation and salinity, with relative contributions being similar among species. GLMs provided ‘outstanding’ model predictions (AUC>0.9) for two of the four fish species while MaxEnt provided ‘excellent’ model predictions (AUC=0.8–0.9) for three of four species. The level of agreement between observed and predicted presence/absence sites for both modelling techniques was ‘moderate’ (K=0.4–0.6) for three of the four species with *P. americanus* models presenting the lowest level of agreement (K<0.1). For the scope of this study, both modelling approaches presented here were determined to produce viable presence/absence maps which represent a snap–shot of the potential distributions of the investigated species. This information provides a better description of demersal fish spatial ecology and can be of a great deal of interest for future fisheries management and conservation planning.

Keywords: demersal fish, Generalized Linear Models, MaxEnt, species distribution models.

Table of Contents

INTRODUCTION	5
MATERIAL AND METHODS	7
Study area	7
Fisheries Surveys and biological data.....	9
Environmental data.....	11
Data Analysis	12
Generalized Linear models.....	12
Maximum entropy model MaxEnt	14
Model performance evaluation	15
RESULTS.....	16
Model fit and explanatory variable contribution.....	16
Comparison of GLM and MaxEnt model performances.....	20
Species habitat suitability maps.....	22
DISCUSSION.....	26
APPENDIX 1.	30
APPENDIX 2	31
REFERENCES	35

INTRODUCTION

Knowledge of geographical distributions patterns of marine species is essential for population's ecology research, fisheries management and biodiversity conservation. Studying the relationship between environmental variables thought to influence species distributions and abundance is crucial to provide useful information for adequate conservation planning and management (Macleod et al., 2008; Franklin, 2009). In this sense, species distribution modelling (SDM) is a mathematical approach used to investigate species–environment interactions by relating occurrence or abundance of species with environmental variables or predictors (Franklin, 2009). With a large variety of modelling techniques combined with the increasing availability of remote sensing data on environmental factors for studying oceans dynamics, SDMs provides the ability to create species potential distribution maps and assess habitat usage and preferences (Guisan & Zimmermann, 2000). SDMs studies for the marine environment have been presented for a broad range of ecological assessments, e.g., analyzing feeding areas of seabirds (Skov et al., 2008), predicting habitat distribution of deep–water corals (Leverette & Metaxas, 2005; Davies & Guinotte, 2011), characterization of macro–epibenthic habitats (Freeman & Rogers, 2003), influence of climate changes on fish distribution (Perry et al., 2005; Lenoir et al., 2011), identification of priority conservation sites in coastal marine environments (Francis et al., 2005) and for implementation of Marine Protected Areas (Conover et al., 2000).

A wide array of modelling techniques have been developing to predict species distribution and detailed comparisons between model performances were presented for various terrestrial species (e.g. Guisan & Zimmermann 2000; Tsoar et al., 2007; Franklin 2009), freshwater fishes (Olden & Jackson, 2002) and marine mammals and fishes (MacLeod et al., 2008, Valavanis et al., 2008; Ready et al., 2010). Most of the modelling techniques require abundance or presence/absence data collected in geographic regions that were systematically surveyed (Guisan & Zimmermann 2000). These techniques vary in how they model the distribution of the response variable, select relevant predictor variables, define fitted functions for each variable, weight variable contributions, allow for interactions, and predict geographic patterns of occurrence or abundance (Elith et al., 2006; Franklin, 2009). Statistical methods used to analyze abundance data include multivariate ordination and clustering analysis,

generalized linear models (GLM), generalized additive models (GAM), classification and regression trees (CART) and multivariate adaptive regression splines (MARS).

Some of these techniques have been extensively used to establish associations between biotic and abiotic environmental features and demersal fish species richness (Leathwick et al., 2006) and relative abundance or other density measurement like CPUE (O'Brien & Rago, 1996; Maravelias et al., 2007a; 2007b; Katsanevakis & Maravelias, 2009), assessed with commercial fishing landings data (Morris & Ball, 2006; Sundermeyer et al., 2006). These studies demonstrated that many of the relationships between fish abundance and environmental variables were shown to be non-linear, indicating that traditional linear methods of statistical analysis (namely multivariate analysis and GLMs) may be inadequate for this kind of analysis because it represents a violation of models assumptions (Maravelias et al., 2003). Additionally, abundance models presented poor performances compared to presence/absence models, mostly due to the fact these models rely on assumptions regarding linearity and symmetric distribution of the response variable (e.g. abundance), that is not in agreement with observations in the field (Francis et al., 2005). Modelling approaches that make use of presence/absence data require accurate data on locations where the species is known to occur (presence data) but also where does not occur (absence data). This will generate statistical functions or discriminative rules allowing habitat suitability to be ranked according to the probability of species presence and absence along environmental gradients (Valavanis et al., 2008). However, this kind of information is not always available, particularly for marine species that occur at great depth and are difficult to detect due to their mobility or inadequate sampling procedures (MacLeod et al., 2008). Even so, a few studies using presence/absence data have been conducted to predict marine fish distributions at a local (Moore et al., 2009; Young et al., 2010), regional (Maravelias et al., 2003; Crec'hriou et al., 2008) and global scale (Ready et al., 2010).

There is a growing interest in models using presence only data, consisting only of observations of the organism presence but with no reliable data on where the species is not found (Pearce & Boyce, 2006). Hence, recent advanced statistical and computer algorithms techniques are now better in recognizing and learning the complex nonlinear relationships between biotic and abiotic aspects of the marine

environment (Phillips et al., 2006; Franklin 2009). Presence only models (e.g. ENFA, BIOCLIM, DOMAIN, GARP, MaxEnt) are adequate for data deficient species and can be used with coarse-scale spatial environmental predictors (see Franklin, 2009). Thus, they were found to be effective in predicting distributions for many species of different regions (Elith et al., 2006).

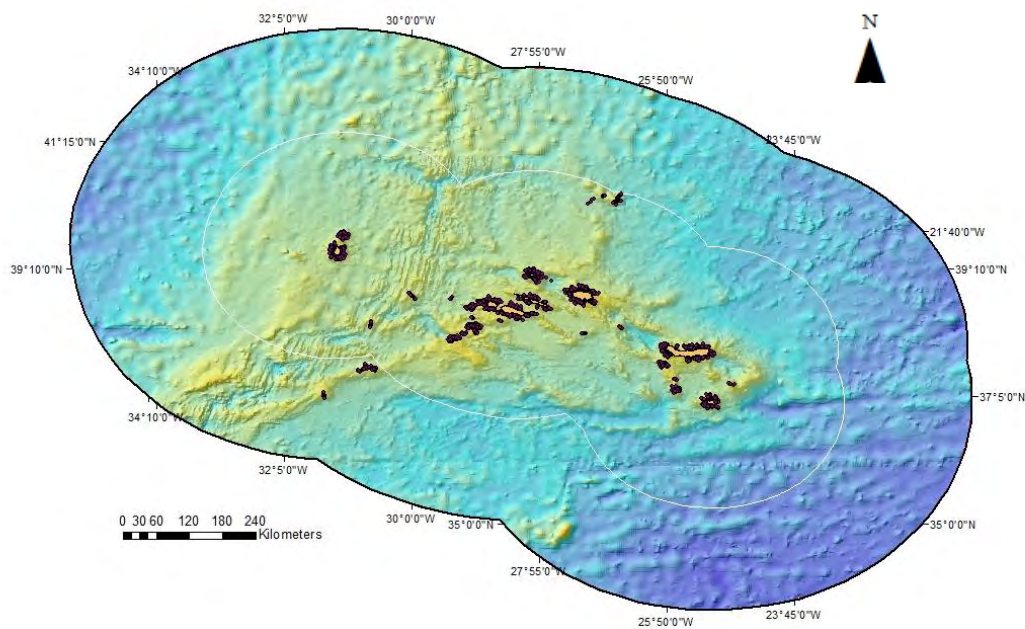
The structure and zonation of demersal fish assemblages in the Azores Archipelago are very well described and relationships between abiotic variables (e.g. mean depth, mean temperature, bottom steepness and irregularity, and depth stratum size) and assemblages composition have been established using multivariate analysis approaches (Menezes et al., 2006; 2009). In the present study we used the maximum entropy model (MaxEnt) and generalized linear models (GLM) to assess the potential distribution in the Azorean economic exclusive zone (EEZ) of four demersal fish species with high commercial and economic interest: blackbelly rosefish, *Helicolenus dactylopterus dactylopterus*, forkbeard, *Phycis phycis*, wreckfish, *Polyprion americanus* and offshore rockfish, *Pontinus kuhlii*. We used fish presence/absence data collected from 13 years of regular longline surveys combined with geostatistical and GIS tools to estimate the probability of occurrence of the target species from seven candidate predictor variables: depth, oxygen saturation, bottom salinity, slope, sediment type, eastness and northness (both derived from aspect). We're interested to determinate the relative importance of the predictors in fish distributions and compare the performance and output of a presence-absence GLM versus presence-only MaxEnt. This study represents the first attempt to apply habitat predictive modelling with environmental grids to predict the distribution of demersal fish species in the Azores EEZ. The potential distribution maps here produced reflect the average distribution of the investigated species and could help in future fisheries management planning.

MATERIAL AND METHODS

Study area

The Azores archipelago is located in the north Atlantic between 36–40°N and 24–32°W and it is composed by 9 volcanic islands distributed in 3 groups along a

tectonic zone extending about 600 km WNW–ESE (Figure 1). The Azorean EEZ comprises a total area of approximately 1.000.000 Km² with an average depth of about 3.000 m. Depths lesser than 600 m comprise a total area of 7.000 Km², less than 1% of the total EZZ area. The seafloor surrounding the Azores islands is characterized by very narrow shelves and steep slopes, and the sea bottom is very irregular and rocky (Menezes, 2003; 2006). Seamounts or submarine elevations are common features in the Azores and may occupy 37 % of the total area of the EEZ (Morato, 2008). They are known to be very important for fisheries and biodiversity (Pitcher et al., 2007).



Fisheries Surveys and biological data

Species investigated in the present study are major components of the Azores bottom longline fisheries and were chosen according to their commercial importance, and homogeneous life history traits and habitat preferences (Table I). These species were: blackbelly rosefish, *Helicolenus dactylopterus dactylopterus*, forkbeard, *Phycis phycis*, wreckfish, *Polyprion americanus* and offshore rockfish, *Pontinus kuhlii*. Species occurrence data were obtained from IMAR/DOP UAz bottom longline surveys performed from 1996 to 2011, with the exception of 1998 and 2006, conducted on coastal and on offshore banks of the Azores archipelago on board of the research vessel “Arquipélago” (Menezes, 2006). A total of 27 cruises comprising 488 bottom longline fishing sets performed within the Azores Economic Exclusive Zone were analyzed.

Table I. Common name, family, habitat and number of presence data points of the four investigated demersal fish species.

Species	Common name	Family	Habitat	Presence data points
<i>Helicolenus dactylopterus</i>	Blackbelly rosefish	Sebastidae	Benthic	3492
<i>Phycis phycis</i>	Forkbeard	Phycidae	Bentopelagic	1249
<i>Polyprion americanus</i>	Wreckfish	Polyprionidae	Benthic	296
<i>Pontinus kuhlii</i>	Offshore rockfish	Scorpaenidae	Benthic	889

Details about the bottom longline surveys sampling design and strategies are presented in Menezes (2003; 2006). Research cruises were completed in spring (393 sets) with few performed in summer (55 sets) and autumn (40 sets). Longline surveys were performed from about 50m to 1200m depth. The fishing gear used in the surveys was similar to the one normally used by Azorean commercial fishery, locally known as ‘stone/buoy bottom longline’ (Figure 2). The longlines are set from four sided skates (each corresponding to a quarter–skate line), with about 30 hooks (hook size n^o 9) by quarter–skate side (each approximately 36.5 m long), baited with ‘chopped salted sardine’. On average 12 skates gear length cover approximately 1 nautical mile. Usually, lines are set at 4:30 AM and hauled at 8:00 AM.

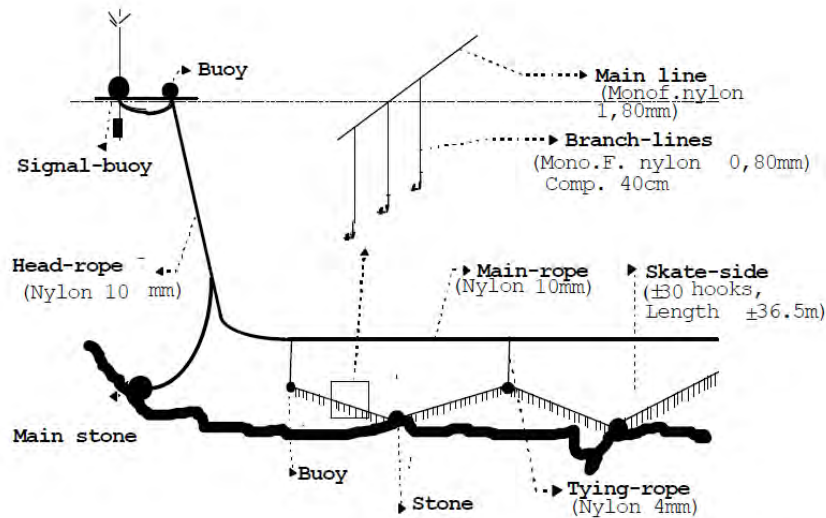


Figure 2. Schematic representation of the bottom longline gear used in the surveys (in Menezes, 2003).

To meet the objectives of this study we started by allocating a geographical position to the presence or absence of fish species in the bottom longlines surveys. Since the position of every fish was not recorded on board the research vessel, we used the best approximation available. The geographic position of the beginning and end of a quarter skate line is recorded when the gear is deployed. However, fish catch was only recorded by 50 m depth strata (Menezes, 2003). To allocate a geographic position to fish presence or absence we 1) estimated the mean geographic position of each quarter skate line and the correspondent depth strata, 2) then analysed the catch by depth strata and allocated fish species presence (1) or absence (0) accordingly. A total of 8124 presence/absence species data sites or locations were obtained.

Geographic positions of quarter skate lines were converted to ESRI shapefile and prepared in ArcMap. From these intermediate positions we kept the ones that represented the beginning and the end of a stratum for each set and connected them by polylines. The longline track polylines were created using the ET GeoWizard “point to polyline” feature in ArcGIS 10 software. For each polyline we calculated the mean longitude and latitude, which represented the geographic position of the species point data to be used in the modeling process. Projected data (WGS 84) were checked and points were removed if they were located outside of the Azorean EEZ or in land.

Environmental data

From a total of twenty one environmental layers of geomorphologic and biogeochemical characteristics available for the Azores EEZ, we choose those that were expected to contribute in explaining fish distribution. These were: depth (m), oxygen saturation (%), bottom salinity (‰), slope (degrees), sediment type (categorical with 7 levels: mud, rock, mixed sediments, sand, muddy sand, coarse sediment and sandy mud), northness and eastness (derived from aspect). All these variables, except for oxygen saturation, have previously been identified as influencing fish distribution (Appendix 1). The inclusion of these variables in the modelling process was dependent on data availability and whether they are known or thought to be directly or indirectly relevant to the target–species distributions. The remaining variables examined (bottom temperature, bathymetric positioning index, slope rugosity, general curvature, plan curvature, profile curvature, bottom pH, bottom alkalinity, bottom Ω aragonite, bottom Ω calcite, phosphate, silicate, nitrate, dissolved oxygen, apparent oxygen utilization) were discarded due to collinearity problems or low contribution to the model construction (data not shown). Details about the environmental layers creation will be presented elsewhere (Tempera, submitted). Fish presence or absence data points in cells with no environmental data values were excluded from further analysis.

Depth values of species occurrences applied in the models were taken from the vessel depth soundings in the surveyed data, known to be more precise than values presented in global depth layers. Oxygen saturation and salinity layers were assembled as long term averages from satellite data, while slope, aspect and sediment type were derived from acoustic bathymetric data. Aspect is a circular rather than a continuous variable and defines the compass orientation in which the slope is facing. We selected this variable as a possible candidate to be included in the modelling process because the aspect of the slope relative to the prevailing ocean current direction may influence habitat suitability and occurrence of the species (Macleod et al., 2008; Monk et al., 2012). Aspect was subsequently converted into eastness and northness to produce two layers where $\text{eastness} = \sin(\text{aspect})$ and $\text{northness} = \cos(\text{aspect})$. These two variables have values varying from -1 to 1 that represent the extent which slope faces north (1), south (-1), and east (1) or west (-1).

We performed a preliminary data exploration analysis in order to avoid multicollinearity between predictors by conducting a Variance Inflation Factor (VIF) analysis and plotting multi-panel scatterplots using BRODGAR 2.5.7. Variables with the highest VIF (>10) were excluded and the remaining data re-analyzed. We repeated this process until all remaining variables had a VIF of less than 10. Also, if a high degree of collinearity exists between two variables (>80%), those variables with the least collinearity with other variables were retained in the modelling process.

Previous to the modelling process, it was necessary to generate two data sets: one presence/absence set for the GLM's and one presence set only for the MaxEnt model. We included all occurrence data available to develop the models, which is hoped to contribute in capturing the species true limits to the tested environmental gradients. Sampling bias can undermine the confidence in species occurrence data, whereas the knowledge of species tolerant limits can be biased by incomplete or limited surveyed data.

For each fish presence or absence data we extracted the length weighted mean values for each continuous environmental variable using the software Geospatial Modelling Environment 0.6.0.0 (Beyer, 2012). Worth to say that fish data are very often autocorrelated in space, time or both (Planque et al., 2011). More particularly, spatial autocorrelation is the tendency of neighboring sample units to possess more similar values than those further apart and represents a potential problem for all area-based studies because the assumption of independency between observations is violated (Fielding & Bell, 1997). There are methods available to face this issue (see Dormann et al., 2007), however we did not further investigated spatial autocorrelation aspects in this study.

Data Analysis

Generalized Linear models

We assessed the relationships between the probabilities of species presence (response variable) and the selected environmental predictors at the sampled locations using binomial generalized linear models (GLM) with a logistic link function (McCullagh & Nelder, 1989). The GLM is a flexible generalization of ordinary least squares regression that allow for non-linearity and non-constant variance structures

in the data, while data can be assumed to be from several probability of distributions (e.g. normal, binomial, gamma, Poisson) which better fit the non-normal error structures of most ecological data (see Guisan et al., 2002). GLMs were fitted using presence-absence data from 8124 observation points and 7 predictor variables in order to select a final model, represented by the best set of predictors that explained species distributions, using the statistical programming environment R v2.14.2 (R Development Core Team, 2012). The selection of significant predictor variables was carried out through manual stepwise forward selection (i.e. process starts with a null model and takes a step by adding one term to the current model) and models were fitted with linear and second order polynomial terms specified for all quantitative predictors. We also plotted density histograms (Appendix 2) showing the distribution of each variable for the presence and absence data of each species to determine if the presence points could be statistically separated from the absence points.

For each target species, final models included predictors which satisfied three criteria: (i) contributed to the reduction of the Akaike's Information Criterion value (AIC; Akaike, 1974), (ii) significant at the 0.05 confidence measured with the chi-square test of deviance reduction, and (iii) had to explain at least 1% of the deviance. For variable selection criteria, AIC is normally used among published SDM studies and it's a measure of the likelihood between the fitted values of the model and the observed values, penalized by the number of parameters in the model (Planque et al., 2011). AIC is based on the principle of parsimony, which helps identifying the model that accounts for the most variation with the fewest variables, and was computed as a measure of the "goodness-of-fit" to determinate variable inclusion or exclusion. However, AIC tends to select models with too many parameters when sample sizes are large (Boyce et al., 2002). Increasing the complexity of a model by including additional terms will increase the accuracy of the regression for the training data but will also tend to decrease the accuracy when it is used for prediction (Venables & Ripley, 2002). Furthermore, *P*-values are considered insufficient (Burnham & Anderson, 2001). Therefore, we decided to adopt Guisan & Hofer (2003) approach by excluding predictors with less than 1 % of deviance reduction. We also followed Maxwell et al. (2009) approach by not including possible interactions between the variables as they were considered to have limited biological meaning. After having selected a final

model, the fitted values were plotted against each selected term in order to characterize the shape of the influence of the predictor on the probability of presence. Finally, we mapped the fitted probabilities of occurrence for each species given by the respective final GLM, using the Marine Geospatial Ecology Toolbox (MGET; Roberts et al., 2010) which integrates ArcGIS with R program language (R Development Core Team, 2012). The GLM tool within the MGET produces a predictive map of species distribution based on the model and input raster layers. Since potential species distributions maps were derived from the model logistic output probabilities of presence, a threshold for considering the species presence had to be adopted (see below).

Maximum entropy model MaxEnt

Presence-only models for each investigated species were developed using MaxEnt software (V. 3.3.3k), downloaded from the MaxEnt software for species habitat modelling web page (www.cs.princeton.edu/~schapire/maxent). This algorithm developed by Phillips et al. (2006), is based on the principles of maximum entropy, whereby a target probability distribution is estimated by finding the probability distribution of maximum entropy, i.e., that is most spread out or closest to uniform, subject to a set of constraints that represent incomplete information about the target distribution. A recent description provided by Elith et al. (2011) stated that MaxEnt minimizes the relative entropy between two probability densities (i.e. one estimated from the presence data and one, from the landscape) defined in covariate space. This technique focuses on fitting a probability distribution of species presence in a set of commonly georeferenced gridded layers or environmental variables, and produce predictions between 0 and 1, representing logistic probabilities of species occurrence (Phillips et al., 2004; 2006). Maxent has been used to study the distribution of both terrestrial (Elith et al., 2006; Peterson et al., 2007) and marine species (Lefkaditou et al., 2008; Ready et al., 2010; Hermosilla et al., 2011; Pittman & Brown, 2011, Jones et al., 2012).

Although MaxEnt is a presence-only algorithm, for the background or absence data we adopted the Elith & Leathwick (2007) approach designated by ‘target-group background’, i.e. localities where other species in the group of interest have been

collected, but not the particular species being modeled. Using observed absences instead of randomly selected pseudo-absences (Philips et al., 2006) was found to improve significantly model performance (Phillips & Dudík, 2008, Mateo et al., 2010). Default settings for features and regularization were used. Data were modeled using the SWD (samples-with-data) format in MaxEnt, i.e. the environmental variables data is presented in spreadsheet-like summaries at both presence and absence (background) sites (Elith et al., 2011). The maximum number of iterations that allow the algorithm to get close to convergence was set to 500. The convergence threshold and regularization multiplier were all left at the default value of 0.0001 and 1 respectively. The relative predictor importance to the model was investigated based on MaxEnt output and the fitted response curve plots of the three most important predictors to the target-species distributions. Since there's no independent data to test models predictions, we used 5 fold cross-validation to get estimates of uncertainty for the response curves. Similar to GLM, the potential species distributions maps were derived from the model output probabilities of presence and a threshold for considering the species presence had to be adopted (see below).

Model performance evaluation

The evaluation of the models was conducted in two parts and focused on the predictive performance at sites. First, the ability of the model to discriminate between presence and absence states was determined by using the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) plot test statistics. AUC has been extensively used in SDM studies. An ROC curve compares the proportion of true positive predictions (sensitivity) and the proportion of false positive predictions (i.e. 1-specificity) obtained from the model across a range of thresholds or cutoff points. AUC ranges from 0 to 1, where a score of 1 indicates perfect discrimination and a score of 0.5 implies predictive discrimination no better than a random guess (Elith et al., 2006). We adopted Hosmer and Lemeshow (2000) interpretation presented in Pittman & Brown (2011), whereby an AUC value of 0.7–0.8 is considered an 'acceptable' prediction, 0.8–0.9 is 'excellent' and >0.9 is 'outstanding'.

However, the sole use of AUC for measuring predictive performance reveals some concern (Austin, 2007; Lobo et al., 2008). Therefore, in the second part of the

evaluation process we used Kappa (Cohen, 1960) statistics, which is a chance-corrected measure of the agreement between observed and predicted values. Fielding & Bell (1997) adapted the classification of Landis & Koch (1977) considered in medical applications and interpreted Kappa values of 0.0–0.4 indicating slight to fair model performance, values of 0.4–0.6 moderate, 0.6–0.8 substantial and 0.8–1.0 almost perfect. K value estimation for GLM and MaxEnt models was performed using the “SDMTools” package in R software (Core Development Team, 2012). Kappa requires a threshold to be applied to the predictions in order to dichotomize the continuous probability output from the models into a binary result (0 or 1), whereby cell values with a probability higher than the selected cutoff value are classified as presence (1), and lower probabilities as absence (0).

For the threshold determination, we decided to use the ROC procedure to identify an optimum probability threshold by reading the point on the curve at which the sum of sensitivity plus specificity is maximized. Liu et al. (2005) found this approach to be relatively good, in which the predictive success of the models is maximized. Thresholds selected in GLM were estimated by the maximization of the Youden index (J , see Perkins & Schisterman, 2006), selected by default within the MGEToolbox, and in MaxEnt models we applied the maximal training sensitivity plus specificity threshold rule. Providing a similar threshold rule in the two modelling techniques will allow direct comparisons between models performances.

RESULTS

Model fit and explanatory variable contribution

The forward stepwise variable selection performed in GLM reduced our environmental predictors set to the following 5 variables: depth, slope, sediment type, sea bottom oxygen saturation and salinity. Although the methods for selecting significant explanatory variables vary between the two modelling approaches used, our results showed that similar explanatory variables were significant for *P. phycis*, *P. americanus* and *P. kuhlii* GLM and MaxEnt models (Table II, III). The most important explanatory variables for *P. Phycis* were depth, slope and oxygen saturation while for *P. americanus* were depth, slope and sediment type. For *P. Kuhlii* the most important explanatory variables were depth, sediment type and slope. For *H. dactylopterus*, the

two modelling techniques selected slightly different significant explanatory variables. Depth and slope were significant for both modelling techniques but salinity was significant in the GLM and oxygen saturation when using MaxEnt modelling technique.

The total deviance values for all GLMs varied from 19.5% for *P. americanus* to 39.6% for *P. phycis*. *Pontinus kuhlii* and *H. dactylopterus* deviance explained was 35.2% and 17.4%, respectively. The explained deviance of the GLM fitted to each species was consistent with those typically fitted for standardization of catch data (Su et al., 2008) and indicated that species distributions were somehow associated with the environmental variables tested in this study. We found that second order polynomial functions gave a better description of environmental associations than linear terms.

For all four species depth was found to be the single most important explanatory variable with relative contributions ranging from 82.1% to 94.9% of the explained deviance for GLMs (Table II) and 79.4% to 94.1% of relative contributions estimated by MaxEnt (Table III), in both case for *P. americanus* and *P. phycis*, respectively. Slope was, in general, the second most important variable explaining the distribution of fish ranging from 2.5% to 13.1% in the GLMs and 2.4% to 5.9 in the MaxEnt. Sediment type and oxygen saturation explanatory power ranged from 2.5% to 8.2% in the GLMs and 1.7% to 6.6% in the MaxEnt.

Table II. The final GLM for each fish species investigated, AIC score, contribution of each significant variable to the percentage of total deviance explained and prevalence (i.e. proportion of sites at which the species were present). 2 alongside the variable indicate the use of second-order polynomial. Osat=Oxygen saturation, Sedt=sediment type.

Species	Fitted parameters	AIC	% of total deviance	Prevalence (%)
<i>H. dactylopterus</i>	Depth ² + Slope ² + Salin	8063	27.4 (82.8 + 13.1 + 4.4)	43
<i>P. phycis</i>	Depth ² + Slope ² + Osat ²	4221	39.6 (94.9 + 2.5 + 2.5)	15
<i>P. americanus</i>	Depth ² + Slope + Sedt	2006	19.5 (82.1 + 2.7 + 8.2)	4
<i>P. kuhlii</i>	Depth ² + Sedt + Slope	3666	35.2 (91.5 + 5.1 + 3.4)	12

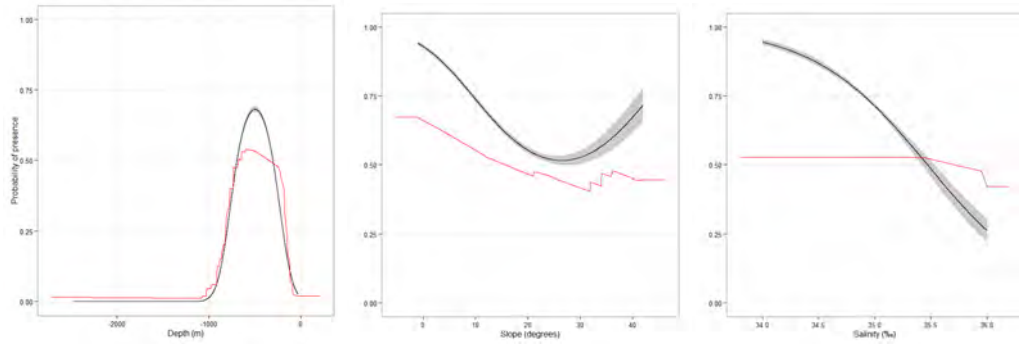
Table III. Relative contribution of the environmental variables estimated by 5 fold cross-validation MaxEnt model for the four demersal fish species investigated.

Species	Variable importance (%)						
	Depth	O ₂ Sat.	Salinity	Slope	Eastness	Northness	Sediment type
<i>H. dactylopterus</i>	89.1	2.7	0.4	5.9	0.2	1.1	0.6
<i>P. phycis</i>	94.1	1.7	0.9	2.4	0.1	0.2	0.5
<i>P. americanus</i>	79.4	4.1	0.7	7.1	1.6	0.5	6.6
<i>P. kuhlii</i>	91.1	1.2	0.1	3	0.4	0.3	3.2

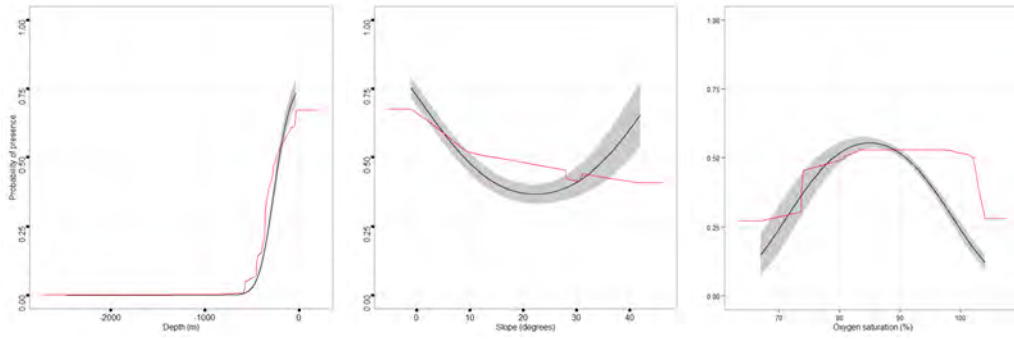
The patterns observed for the response curves of environmental predictors were similar despite the modelling technique used (Figure 3). This is an important result leading to an increased degree of consistency and confidence in the models developed. Additional insights to the underlying ecology of the studied species can be drawn from the fitted response curves. Worth to mention that fitted probabilities of occurrence do not describe the physiological limits of the species, but can be seen as graphical descriptions of how the environmental variables included in the models correlate with the empirical distribution (Chatfield et al., 2010). Nevertheless, the following analyses highlight some broad patterns of the modelled species distribution.

Our models showed a wider bathymetric distribution for *H. dactylopterus*, ranging from 84 to 1000 m depth, with increasing probabilities of occurrence from shallow than 500 m and decreasing for greater depths. In contrast, *P. phycis* showed a narrower bathymetric distribution, with decreasing probabilities of occurrence from 30 to 500 m depth. Regarding slope, all four species presented higher probabilities of occurrence in flat areas or steep slopes, but with *P. americanus* showing a decrease in suitable habitat as the slope increased. Sediment type was particularly important for *P. americanus* and *P. kuhlii*, where both showed a greater association with sandy mud habitats, while *P. americanus* and *P. kuhlii* showed stronger associations with rock and sandy bottoms, respectively. Oxygen saturation was the third most important predictor for *P. phycis* and *H. dactylopterus* MaxEnt model, which illustrated similar response curves but with increased probabilities of occurrence on locations with middle oxygen saturation values. Salinity was important for *H. dactylopterus* distribution, where this species is likely to be found in sites close to 35 ‰. The other variables tested in this study showed no clear response or association with the target-species distributions.

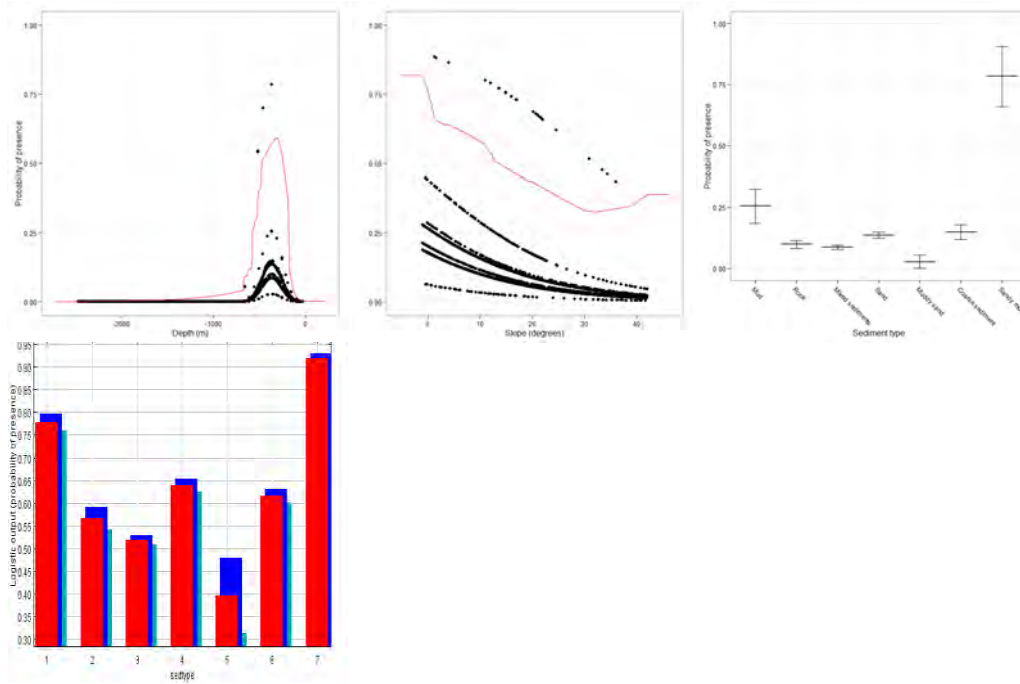
Helicolenus dactylopterus (Blackbelly rosefish)



Phycis phycis (Forkbeard)



Polyprion americanus (Wreckfish)



Pontinus kuhlii (Offshore rockfish)

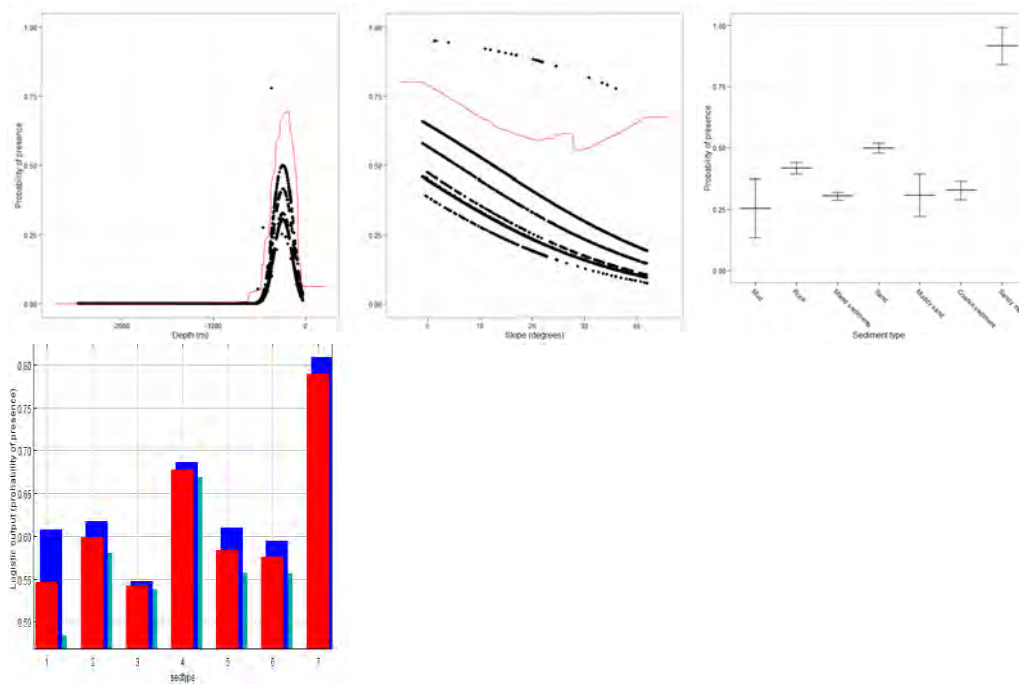


Figure 3. Response curves for MaxEnt and GLM analysis relating target species fitted probabilities of occurrence to the three most important environmental predictors. GLM fitted probabilities were plotted by modelling the original sample data (black circles and lines) with the final model along with estimated standard error (grey). MaxEnt generated response curve plots represent the mean response of 5-fold cross validation replicates (red) and the mean \pm one standard deviation (blue) for the sediment type variable. Sediment type classes were: 1, mud; 2, rock; 3, mixed sediments; 4, sand; 5, muddy sand; 6, coarse sediment; and 7, sandy mud.

Comparison of GLM and MaxEnt model performances

Our results indicate that the models developed here can be used to predict the likely distributions of the investigated species. According to Hosmer & Lemeshow (2000) AUC interpretation, all of the 8 models developed here (two for each of the four species) had at least ‘acceptable’ discriminatory ability ($AUC > 0.7$) and 2 were ‘outstanding’ ($AUC > 0.9$). GLM showed slightly higher AUC scores than MaxEnt for all of the four species. *P. phycis* and *P. kuhlii* GLMs provided ‘outstanding’ predictions ($AUC > 0.9$) and the remaining two species ‘excellent’ ($AUC = 0.8–0.9$). MaxEnt models provided ‘excellent’ predictions for three species, and ‘acceptable’ predictions ($AUC = 0.7–0.8$) for *H. dactylopterus*. Threshold probability values determined for each model technique are not directly comparable but showed some agreement between 3

of the four modelled species, with values ranging from 0.032 to 0.425 in GLM and 0.298 to 0.431 in MaxEnt (Table IV).

Modelling performances evaluated by Kappa analysis showed similar results between the two modelling techniques. However the highest performing model measured by AUC did not have the highest Kappa score. Kappa scores ranged from 0.064 to 0.486, with the highest score attributed to GLM for three of the four investigated species. Following Kappa interpretation offered by Landis & Koch (1977), two species presented moderate (K=0.4–0.6) model performance while the remaining two species provided slight to fair (K=0.0–0.4) model performance. Kappa estimates showed *H. dactylopterus* GLM and *P. phycis* MaxEnt models to be the most accurate, with scores of 0.486 and 0.464, respectively (Table IV). In contrast, the worst accuracy performance was obtained for *P. americanus* GLM and MaxEnt models with K values of 0.101 and 0.064, respectively.

GLM also presented higher rates of true positive predictions (sensitivity) for all species, ranging from 86% to 93% while in MaxEnt models varied from 65% to 83% (Table IV). True negatives or absence predictions (specificity) rates were not so clear in distinguishing the best modelling approach considered here. MaxEnt specificity rates ranged from 57% to 87% presenting the highest values in two species and the lowest for one of the four species, while GLM specificity rates varied from 63% to 79% (Table IV). *P. phycis* and *P. americanus* GLMs presented the highest sensitivity rates, with 93.2 and 91.2%, respectively, while *P. phycis* MaxEnt and *P. kuhlii* GLM and MaxEnt showed the highest specificity rate, with 87.4 and 79.1%, respectively (Table IV).

Table IV Summary of model predictive performances for each investigated species measured using the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC), Cohen’s Kappa statistics (K), sensitivity (% correctly predicted presences) and specificity (% correct predicted absences) calculated from the training data. Presences and absences for assessing Kappa scores, sensitivity and specificity rates were determined using the selected optimum threshold.

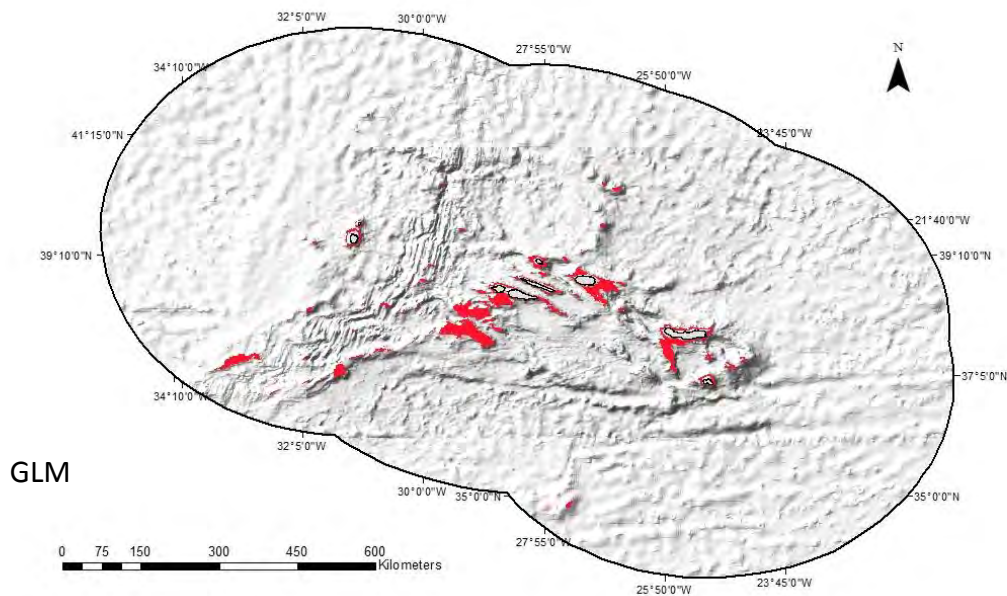
Species	Model	AUC	Threshold	K	Sensitivity(%)	Specificity(%)
<i>H. dactylopterus</i>	GLM	0.822	0.425	0.486	86.7	63.9
	MaxEnt	0.707	0.431	0.453	74.3	71.6
<i>P. phycis</i>	GLM	0.909	0.138	0.457	93.2	75.9
	MaxEnt	0.856	0.397	0.464	65.7	87.4
<i>P. americanus</i>	GLM	0.842	0.032	0.101	91.2	64.4
	MaxEnt	0.823	0.304	0.064	83.7	57.2

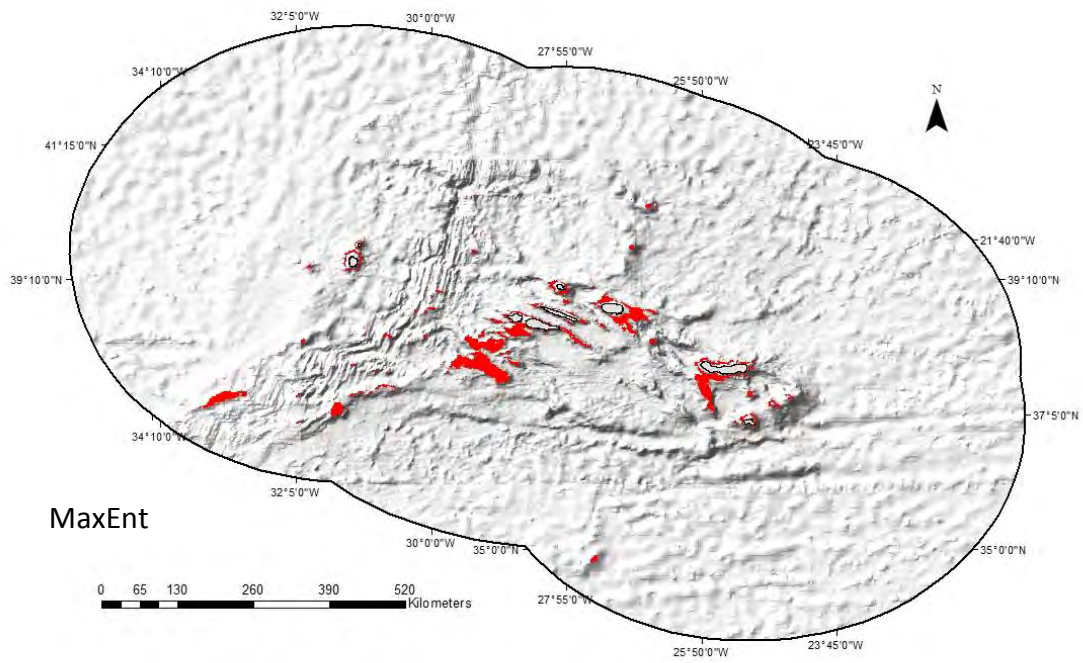
<i>P. kuhlii</i>	GLM	0.905	0.165	0.396	88.0	79.1
	MaxEnt	0.870	0.298	0.336	76.0	79.1

Species habitat suitability maps

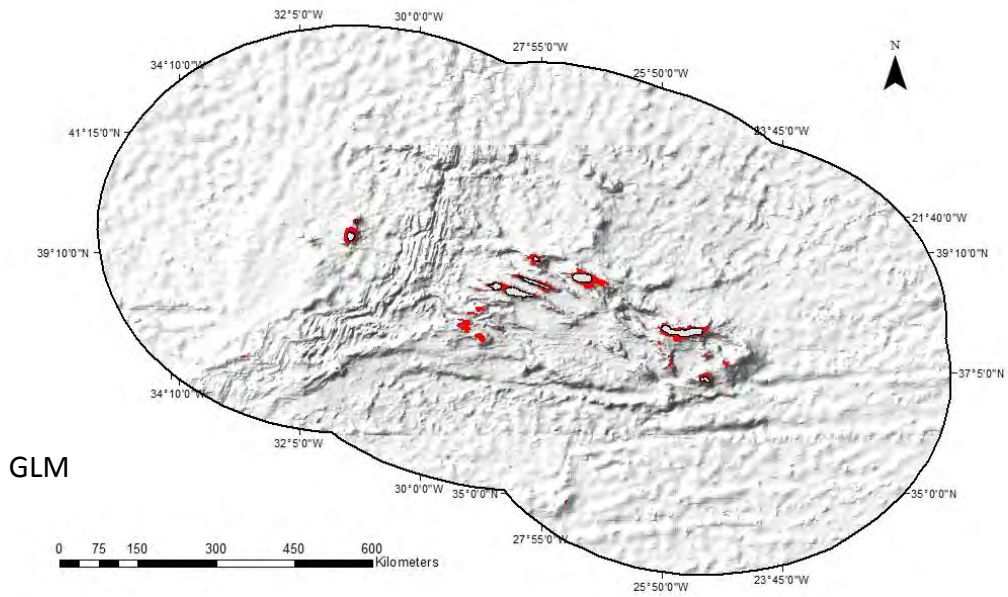
The final habitat predictive models for all species that resulted from GLMs or MaxEnt were extremely similar (Figure 4). After applying the threshold value (Table IV) to the predicted probability of occurrence for each species, we found that all of the four modelled species occur on the islands slopes and offshore banks in the Azores EEZ, with different amplitudes of predicted suitable habitat. *H. dactylopterus* showed a wider distribution along the banks and on less steep slopes around the islands (Figure 4a), while *P. phycis* and *P. kuhlii* habitat suitability map revealed a more restricted distribution compared to the other species modelled (Figure 4b and d, respectively). *P. americanus* distribution map presented a wider suitable habitat areas on the offshore banks and narrower near the islands slopes (Figure 4c).

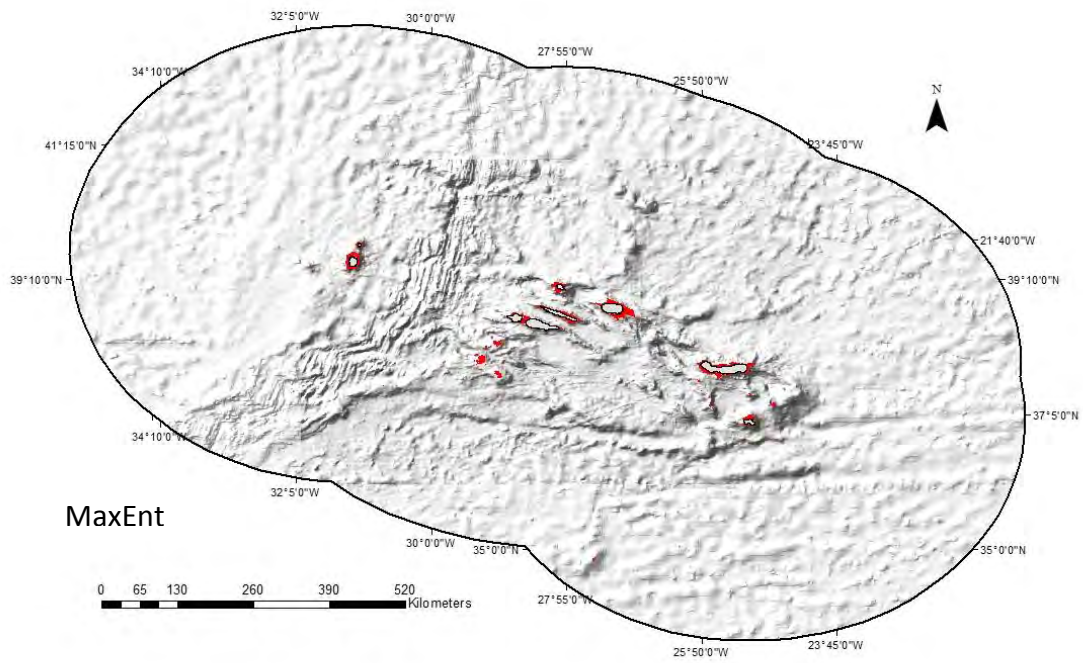
(a) *Helicolenus dactylopterus*



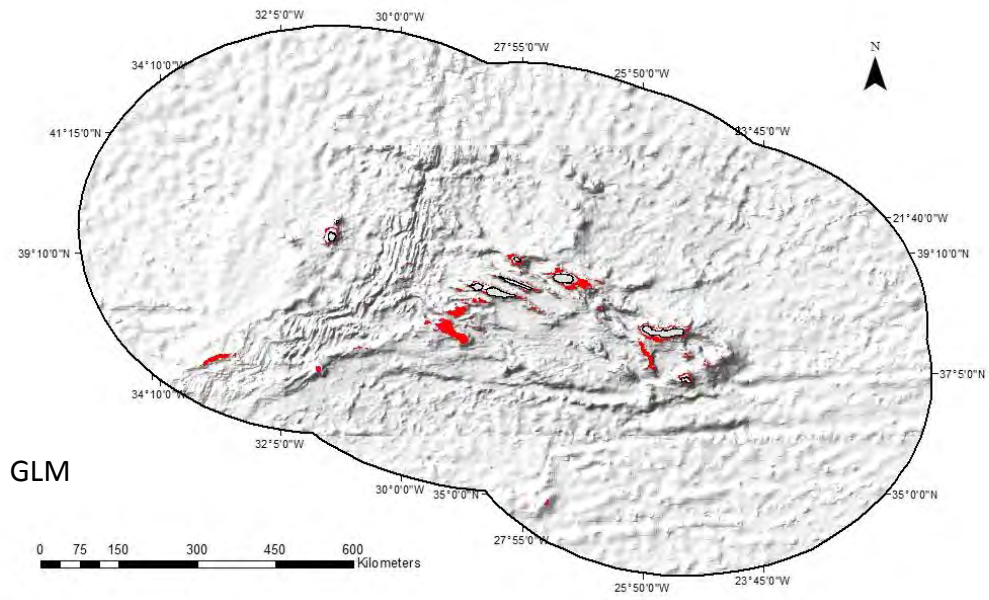


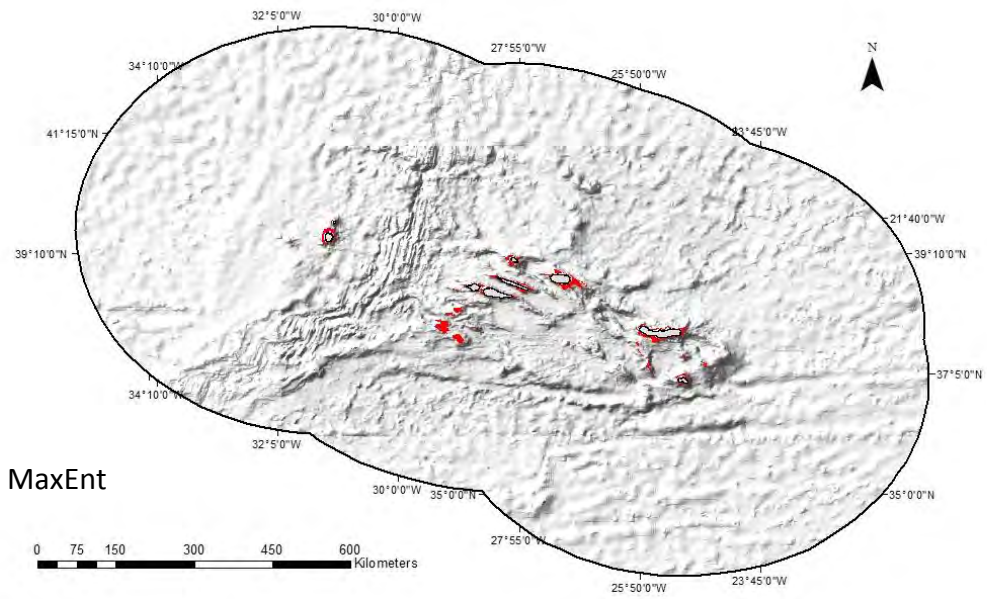
(b) *Phycis phycis*



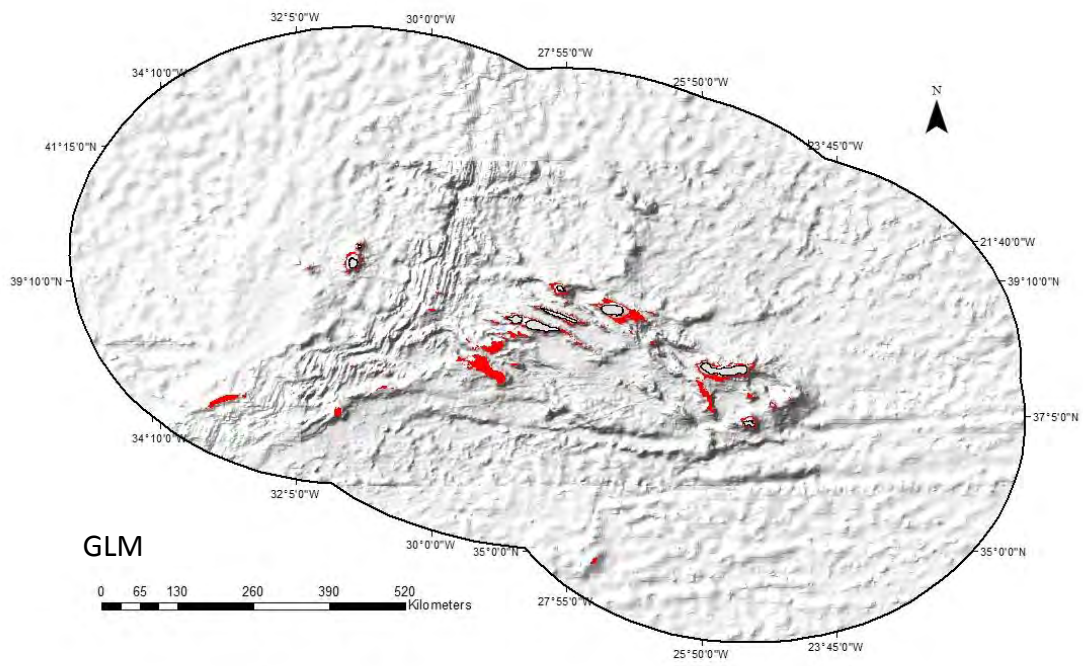


(c) *Polyprion americanus*





(d) *Pontinus kuhlii*



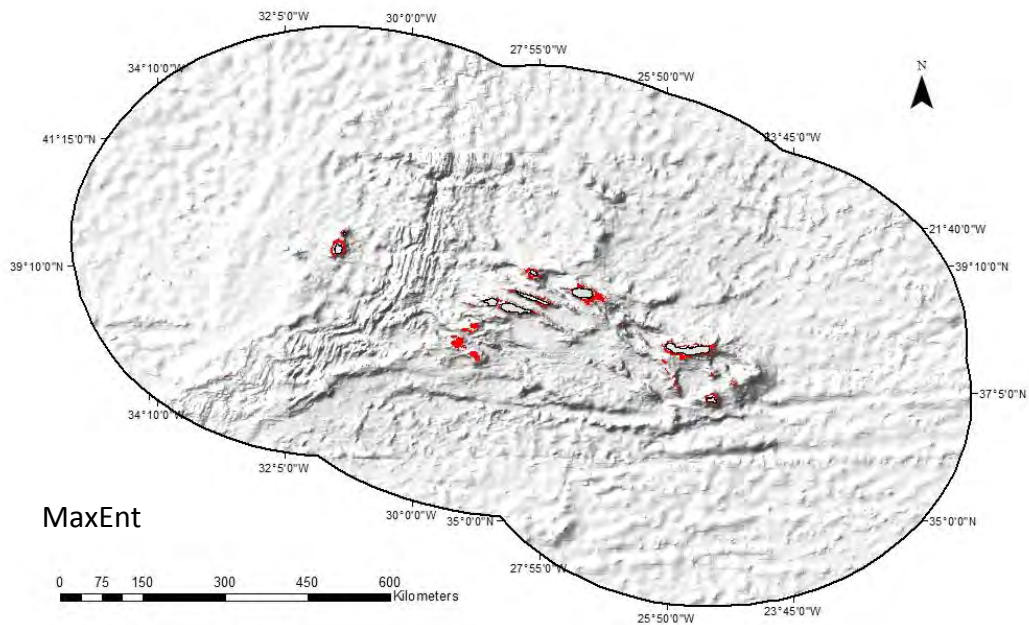


Figure 4. Predicted distributions defined by the GLMs and MaxEnt models for (a) *H. dactylopterus*, (b) *P. phycis*, (c) *P. americanus* and (d) *P. kuhlii*. Predicted presence areas (in red) were estimated using the optimum thresholds (see Table IV).

DISCUSSION

Our study explored the application of MaxEnt and GLMs to high resolution seafloor topographic and biogeochemical data to accurately predict the potential distributions of four economically important demersal fish species in the Azores EEZ. In general, we found species predicted spatial distributions restricted to habitats over seamounts, offshore banks and islands slopes, but with different suitable habitat ranges among the target species. Our models showed relative proportions of potential suitable habitat to the entire Azores EEZ of 1,2% for blackbelly rosefish, 0,5% for forkbeard and wreckfish and 0.3% for offshore rockfish, which comprises total areas ranging approximately from 11.187 to 3.070 Km².

The models developed here contained the significant predictors depth, sea bottom slope, sediment type, oxygen saturation and salinity. Depth was found to be the most important predictor for all species because it was the most detailed layer

used in this study. In contrast, sea bottom oxygen saturation and salinity layers tested in our study were derived from global data sets and perhaps do not reflect the true ranges of these variables in the study area. The effect of sea bottom salinity in defining demersal fish habitat is not so clear with studies demonstrating to be determinant (Maravelias et al., 2007a) while others showed that it did not appear to explain fish spatial distribution (Lauria et al., 2011). Beside the marginal effect in species occurrences caused by the later predictors, the models emphasized the importance of depth and bottom slope in determining the target species distributions. Many other SDM studies have demonstrated the influence of these variables on demersal fish assemblage structure (García–Charton & Pérez–Ruzafa, 2001; Menezes, 2003; 2006; Moore et al., 2009), occurrence (Maravelias et al., 2003; Crec’hriou et al., 2008; Maxwell et al., 2009; Chatfield et al., 2010; Young et al., 2010; Ready et al., 2011) and relative abundance (Morris & Ball, 2006, Sundermeyer et al., 2006; Maravelias et al., 2007b; 2007c; Katsanevakis & Maravelias, 2009; Lauria et al., 2011).

Our GLM analysis indicated that 19.5 to 39.6% of the variation in the target species distributions could be explained by the environmental variables included in the final models, confirming their importance in dictating species distributions. The remaining 80.5 to 60.4% of unexplained deviance for the four species can be attributed to unmeasured environmental variables found important in influencing fish distributions, such as temperature (Perry et al., 2005; Maravelias et al., 2007b) and exposure to currents or wave action (Fulton and Bellwood, 2004). Other causes of unexplained deviance may be attributed to unaccounted for ecological characteristics, life history, and behavioural traits such as diet, size, mobility, predation, and competition (Guisan & Zimmermann, 2000; Franklin 2009; Moore et al, 2009; Chatfield et al., 2010; Young et al., 2010).

Previous researches have addressed the problem of comparing distinct modelling techniques, namely with uncertainties associated with the data–type used, model parameterization processes and underlined assumptions behind each modelling mechanism (Jones et al., 2012). While direct comparisons can be questioned, establishing robust methods to understand and define marine species distributions is crucial to effective fisheries management and conservation planning (Moore et al., 2009). All models developed here had good discriminatory abilities (AUC >0.5).

The lowest AUC score was reported for *H. dactylopterus* MaxEnt model which can be attributed to the fact that this species presented a wider spatial distribution compared to the other species modelled. A wider niche corresponds generally with a lower AUC values (Phillips et al., 2004). In contrast, only three of the four modelled species provided 'moderate' agreement between observed and predicted values assessed by Kappa statistics. *P. americanus* MaxEnt and GLM models had consistently low Kappa scores compared to the other species investigated, resulting in a low specificity rate, although with a relatively high sensitivity rate. We believe that the low number of presence observations (prevalence) in the training data set for this species influenced model performance.

The effect of prevalence on performance measures has been well documented (e.g. Fielding & Bell 1997; Manel et al., 2001). AUC values are independent of the prevalence of the organism being measured and our results are in accordance with this statement, with examples of high AUC scores calculated for both high and low prevalence species. However, the effect of prevalence on Kappa is not so clear. Manel et al. (2001) in their work found that the effect of prevalence appear to be negligible, but other study in a different biological field showed some concerns that kappa is affected by low prevalence (Ridenour & Heath 1999).

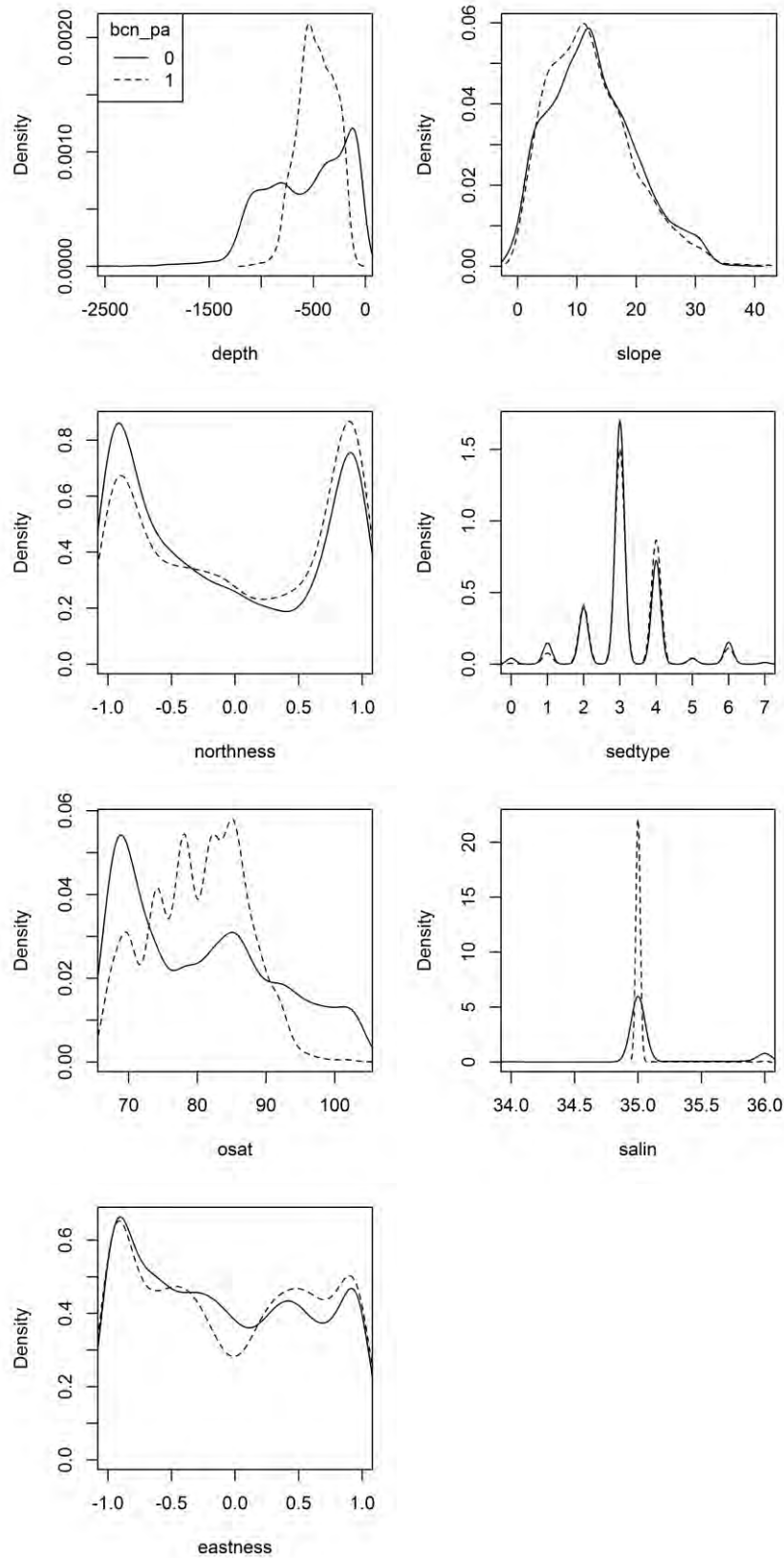
Furthermore, the optimum threshold estimated for the production of the binary fish presence absence distribution maps differed between modelling techniques for three of the four species. Liu et al. (2005) suggested that taking the prevalence of model building data as a threshold helps in obtain a good presence/absence prediction. The optimum threshold probability estimated for the GLMs for each species were very close to the prevalence which they had on the training data set, and consequently these models revealed higher sensitivity rates. In contrast, the estimated specificity rates were not so clear in distinguishing the most accurate technique, with MaxEnt models showing the highest and the lowest rates for *P. phycis* and *P. americanus*. Models performances can be degraded due to the similarity of negative locations to positive locations, resulting in too many false presences and/or absences. There are a variety of ecological processes, operating over a range of timescales that can give rise to data of this type (see Fielding & Bell, 1997). Overall, the predictive ability of both methods was very similar across all but the lowest-prevalence species.

The assessment provided by this work is hampered by the lack of an appropriate spatial autocorrelation analysis. Nevertheless, our use of different modelling techniques and performance measures combined with potential distribution maps, provided insights to species–environment relationships and a better understanding of demersal fish species spatial ecology in the Azores region. Our range of predictions for these four fish species should be subject to future testing by the collection of new data, especially in predicted areas that were previously unsampled. Further analysis in the future will include abundance data in order to assess fish preferred habitat and to explore temporal variations in habitat selection. Thus, more robust models should incorporate fishing effort data in order to quantify how the predicted potential occurrence areas for demersal fish species are affected by fishing pressure and help to define closure areas for stocks recovering.

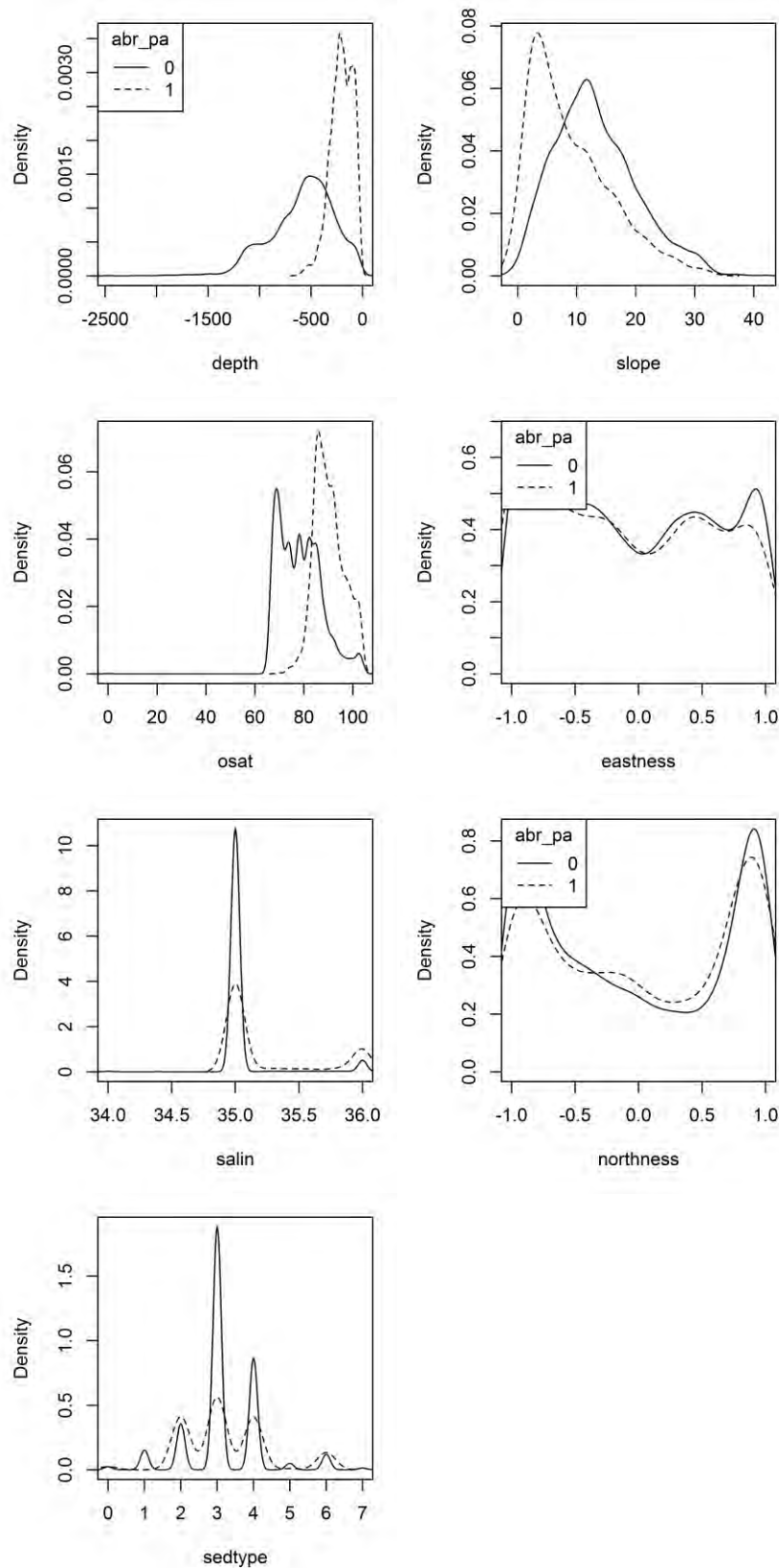
APPENDIX 1. Examples of previous research that have investigated the influence that physical and biological variables have on patterns of fish distribution and relative abundance.

Environmental variable	Study
Water depth	Maravelias et al., 2003; Morris & Ball, 2006; Maravelias et al., 2007a,b and c; Vaz et al., 2008; Maxwell et al., 2009; Moore et al., 2009; Chatfield et al., 2010; Ready et al., 2010; Young et al., 2010; Lauria et al., 2011; Lenoir et al., 2011
Slope	Anderson et al., 2009; Moore et al., 2009; Young et al., 2010
Sediment and substrate type	Morris & Ball, 2006; Sundermeyer et al., 2006; Vaz et al., 2008; Maxwell et al., 2009; Chatfield et al., 2010; Lauria et al., 2011
Bottom Salinity	Maravelias et al., 2007a; Lauria et al., 2011
Northness, eastness and parameters derived from aspect	Pittman & Brown, 2011; Monk et al 2012

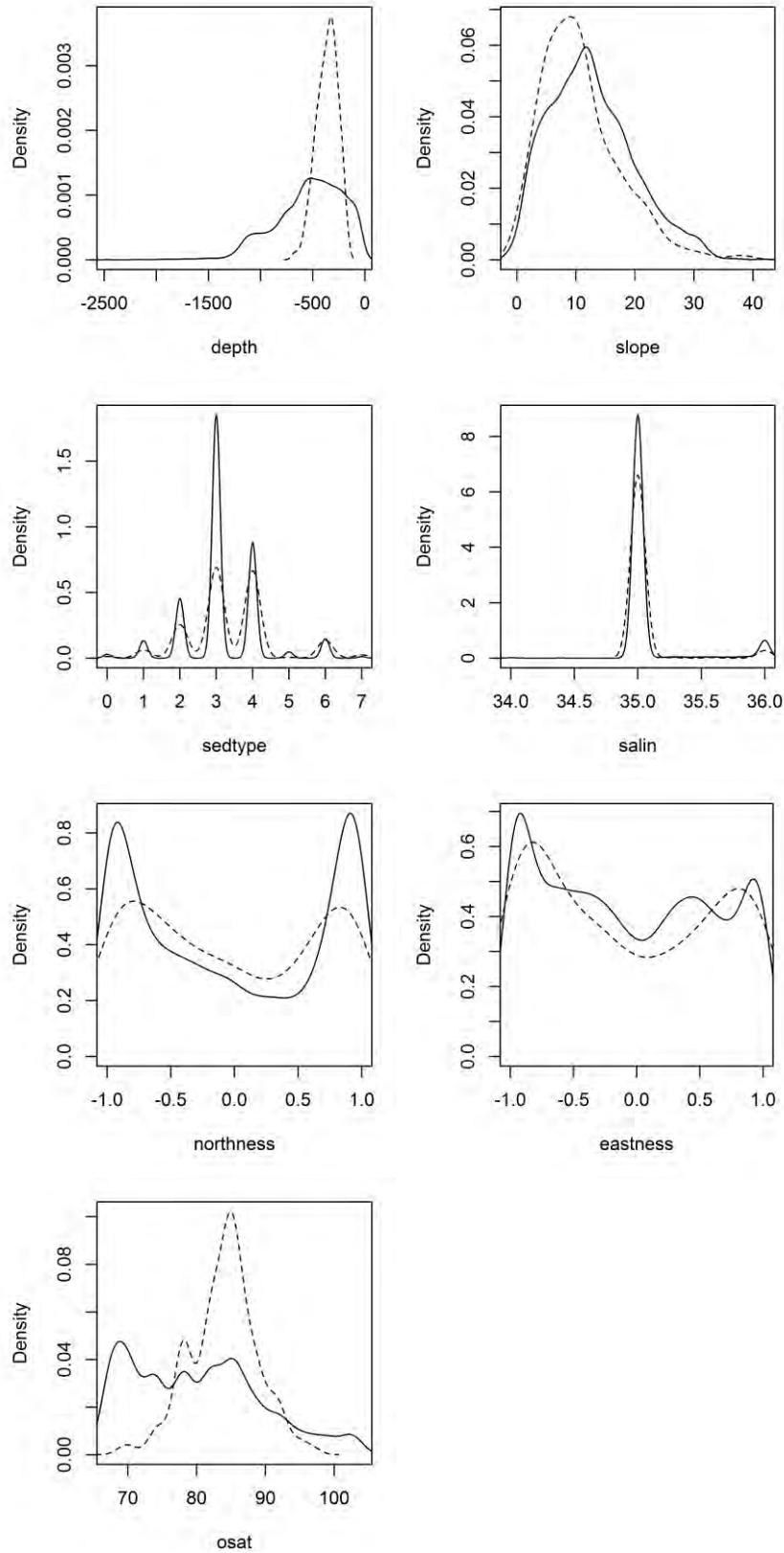
APPENDIX 2. Density histograms of *Helicolenus dactylopterus* presence (dashed line) and absence (solid line) points along with the seven candidate predictor variables. Sediment type classes were: 1, mud; 2, rock; 3, mixed sediments; 4, sand; 5, muddy sand; 6, coarse sediment; and 7, sandy mud.



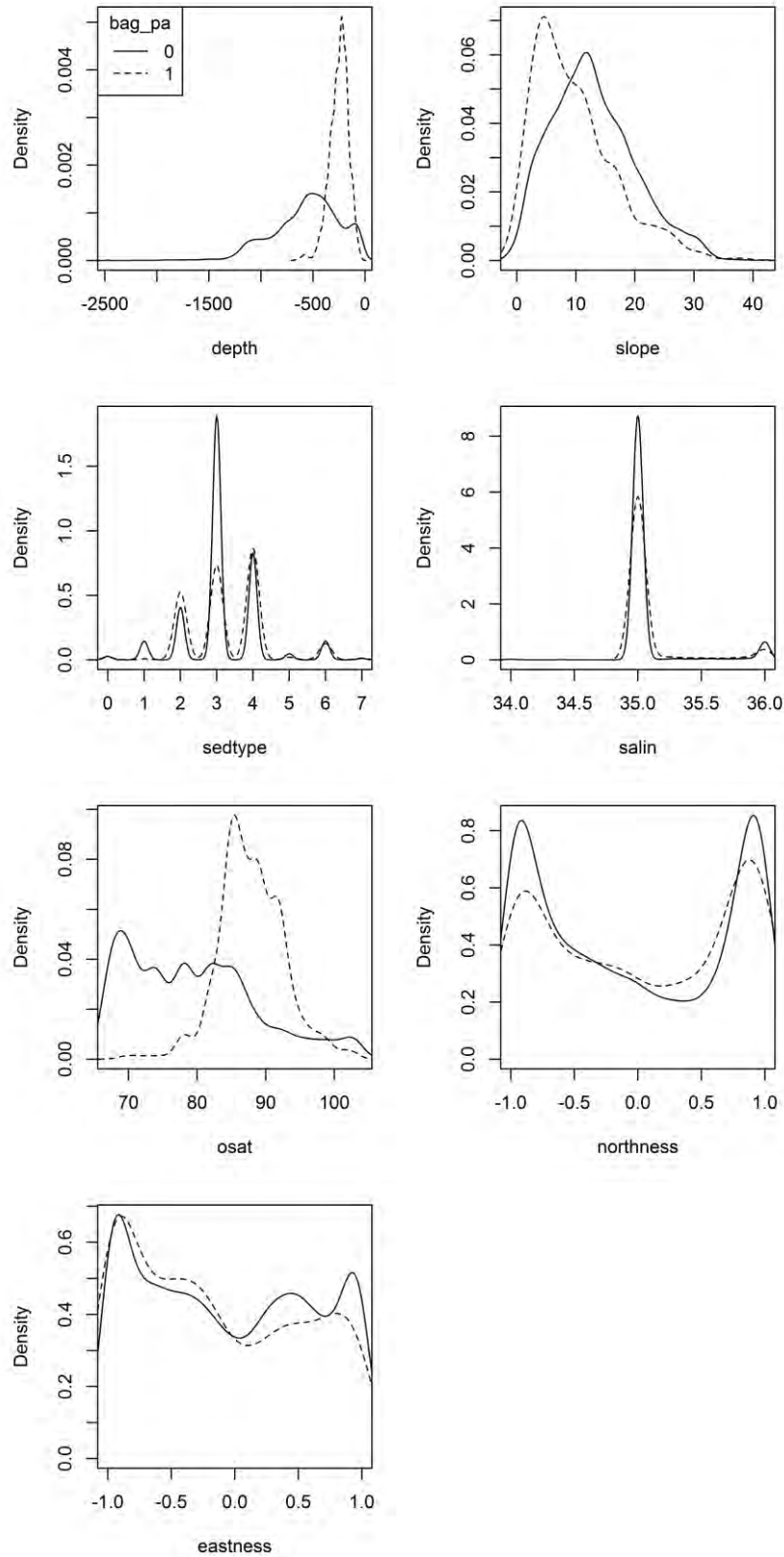
Density histograms of *Phycis phycis* presence (dashed line) and absence (solid line) points along with the seven candidate predictor variables. Sediment type classes were: 1, mud; 2, rock; 3, mixed sediments; 4, sand; 5, muddy sand; 6, coarse sediment; and 7, sandy mud.



Density histograms of *Polyprion americanus* presence (dashed line) and absence (solid line) points along with the seven candidate predictor variables. Sediment type classes were: 1, mud; 2, rock; 3, mixed sediments; 4, sand; 5, muddy sand; 6, coarse sediment; and 7, sandy mud.



Density histograms of *Pontinus kuhlii* presence (dashed line) and absence (solid line) points along with the seven candidate predictor variables. Sediment type classes were: 1, mud; 2, rock; 3, mixed sediments; 4, sand; 5, muddy sand; 6, coarse sediment; and 7, sandy mud.



REFERENCES

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.
- Austin, M.P., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, 200:1–19.
- Beyer, H.L., 2012. Geospatial Modelling Environment (Version 0.6.0.0), (software). URL: <http://www.spatalecolony.com/gme>.
- Boyce, M.S., P.R. Vernier, S.E. Nielsen & F.K.A. Schmiegelow, 2002. Evaluating resource selection functions. *Ecological Modelling*, 157:281–300.
- Burnham, K.P. & D.R. Anderson, 2001. Kullback–Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, 28:111–119.
- Chatfield, B.S., K.V. Niel, G.A. Kendrick & E.S. Harvey, 2010. Combining environmental gradients to explain and predict the structure of demersal fish distributions. *Journal of Biogeography*, 37:593–605.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational Psychology Measurements*, 20:37–46.
- Conover, D., J. Travis & F.C. Coleman, 2000. Essential fish habitat and marine reserves: an introduction to the Second Mote Symposium in Fisheries Ecology. *Bulletin of Marine Science*, 66:527–534.
- Crec'hriou, R., P. Bonhomme, G. Criquet, G. Cadiou, P. Lenfant, G. Bernard, E. Roussel, L.L. Diréach, & S. Planes, 2008. Spatial patterns and GIS habitat modelling of fish in two French Mediterranean coastal areas. *Hydrobiologia*, 612:135–153.
- Davies, A.J. & J.M. Guinotte, 2011. Global habitat suitability for framework-forming cold-water corals. *PLoS ONE* 6(4): e18483. doi:10.1371/journal.pone.0018483.
- Dormann, C.F., J.M. McPherson, M.B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. Davies, A. Hirzel, W. Jetz, W. Kissling, I. Kuhn, R. Ohlemuller, P. Peres-Neto, B. Reineking, B. Schroder, F. Schurr & R. Wilson, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30:609–628.
- Elith, J., C.H. Graham, R.P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J.M. Overton, A.T. Peterson, S.J.

- Philips, K. Richardson, R. Scachetti–Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz & N.E. Zimmermann. 2006. Novel methods improve predictions of species' distributions from occurrence data. *Ecography*, 29:129–151.
- Elith, J. & J. Leathwick, 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity Distribution*, 13:265–275.
- Elith, J., S.J. Phillips, T. Hastie, M. Dudik, Y.E. Chee & C.J. Yates, 2011. A statistical explanation of MaxEnt for ecologists. *Diversity Distribution*, 17:43–57.
- Fielding, A.H. & J.F. Bell, 2007. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24:38–49.
- Francis, M.P., M.A. Morrison, J. Leathwick, C. Walsh & C. Middleton, 2005. Predictive models of small fish presence and abundance in northern New Zealand harbours. *Estuarine, Coastal and Shelf Science*, 64:419–435.
- Franklin, J., 2009. Mapping species distributions. *Cambridge University Press* pp 278.
- Freeman, S.M. & S.I. Rogers, 2003. A new analytical approach to the characterization of macro–epibenthic habitats: linking species to the environment. *Estuarine, Coastal and Shelf Science*, 56:749–764.
- Fulton, C.J. & D.R. Bellwood, 2004. Wave exposure, swimming performance, and the structure of tropical and temperate reef fish assemblages. *Marine Biology*, 144:429–437.
- García–Charton, J.A. & Á. Pérez–Ruzafa, 2001. Spatial pattern and the habitat structure of a Mediterranean rocky reef fish local assemblage. *Marine Biology*, 138:917–934.
- Guisan, A., & N.E. Zimmermann, 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135:147–186.
- Guisan, A. & U. Hofer, 2003. Predicting reptile distribution at the mesoscale: relation to climate and topography. *Journal of Biogeography*, 30:1233–1243.
- Guisan, A., T.C. Edwards, Jr. & T. Hastie, 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157:89–100.
- Hermosilla, C., F. Rocha & V.D. Valavanis, 2011. Assessing *Octopus vulgaris* distribution using presence–only model methods. *Hydrobiologia*, 670:35–47.

- Hosmer, D.W. & S. Lemeshow, 2000. Applied Logistic Regression, second edition. New York, USA: John Wiley & Sons.
- Jones, M.C., S.R. Dye, J.K. Pinnegar, R. Warren & W.W. Cheung, 2012. Modelling commercial fish distributions: Prediction and assessment using different approaches. *Ecological Modelling*, 225:133–145.
- Katsanevakis, S. & C.D. Maravelias, 2009. Bathymetric distribution of demersal fish in the Aegean and Ionian Seas based on generalized additive modeling. *Fish Science*, 75:13–23.
- Landis, J.R. & G.C. Koch, 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33: 159–74.
- Lauria, V., S. Vaz, C.S. Martin, S. Mackinson & A. Carpentier, 2011. What influences European plaice (*Pleuronectes platessa*) distribution in the eastern English Channel? Using habitat modelling and GIS to predict habitat utilization. *ICES Journal of Marine Science*, 68:1500–1510.
- Leathwick, J.R., J. Elith, M.P. Francis, T. Hastie & P. Taylor, 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology–Progress Series*, 321:267–281.
- Lefkaditou, E., C. Politou, A. Palialexis, E.J. Dokos, P. Cosmopoulos & V.D. Valavanis, 2008. Influences of environmental variability on the population structure and distribution patterns of the short-fin squid *Illex coindetii* (Cephalopoda: Ommastrephidae) in the Eastern Ionian Sea. *Hydrobiologia*, 612:71–90.
- Lenoir, S., G. Beaugrand & E. Lecuyer, 2011. Modelled spatial distribution of marine fish and projected modifications in the North Atlantic Ocean. *Global Change Biology*, 17:115–129.
- Leverette, T.L. & A. Metaxas, 2005. Predicting habitat for two species of deep-water coral on the Canadian Atlantic continental shelf and slope. In: Freiwald, A., J. M. Roberts (eds), Cold-water Corals and Ecosystems, pp 467–479.
- Liu, C., P.M. Berry, T.P. Dawson & R.G. Pearson, 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28:385–393.
- Lobo, J.M., A. Jiménez-Valverde & R. Real, 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17:145–151.
- MacLeod, C. D., L. Mandleberg, C. Schweder, S.M. Bannon & G.J. Pierce, 2008. A comparison of approaches for modelling the occurrence of marine animals. *Hydrobiologia*, 612:21–32.

- Manel, S., H.C. Williams & S.J. Ormerod, 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38:921–931.
- Maravelias, C.D., J. Haralabous & C. Papaconstantinou, 2003. Predicting demersal fish species distributions in the Mediterranean Sea using artificial neural networks. *Marine Ecology Progress Series*, 255:249–258.
- Maravelias, C.D., E.V. Tsitsika & C. Papaconstantinou, 2007a. Evidence of Morocco dentex (*Dentex maroccanus*) distribution in the NE Mediterranean and relationships with environmental factors determined by Generalized Additive Modelling. *Fisheries Oceanography*, 16:294–302.
- Maravelias, C.D., E.V. Tsitsika & C. Papaconstantinou, 2007b. Seasonal Dynamics, environmental preferences and habitat selection of John Dory (*Zeus faber*). *Estuarine, Coastal and Shelf Science*, 72:703–710.
- Maravelias, C.D., E.V. Tsitsika & C. Papaconstantinou, 2007c. Environmental influences on the spatial distribution of European hake (*Merluccius merluccius*) and red mullet (*Mullus barbatus*) in the Mediterranean. *Ecological Research*, 22:678–685.
- Mateo, R.G., T.B. Croat, Á.M. Felicísimo & J. Muñoz, 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo–absences and target–group absences from natural history collections. *Diversity Distribution*, 16:84–94.
- Maxwell, D.L., V. Stelzenmüller, P.D. Eastwood & S.I. Rogers, 2009. Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*) and thornback ray (*Raja clavata*) in UK waters for marine management and planning. *Journal of Sea Research*, 61:258–267.
- McCullagh, P. & J.A. Nelder, 1989. Generalized Linear Models. Chapman and Hall, London.
- Menezes, G.M., 2003. Demersal fish assemblages in the Atlantic Archipelagos of the Azores, Madeira, and Cape Verde. PhD thesis, Universidade dos Açores, Horta.
- Menezes, G.M., M.F. Sigler, H.M. Silva, & M.R. Pinho, 2006. Structure and zonation of demersal fish assemblages off the Azores Archipelago (mid–Atlantic). *Marine Ecology Progress Series*, 324:241–260.
- Menezes G.M., A. Rosa, O. Melo & M.R. Pinho, 2009. Demersal fish assemblages off the Seine and Sedlo seamounts (northeast Atlantic). *Deep sea Research*, 56(25):2683–2704.

- Monk, J., D. Ierodiaconou, E. Harvey, A. Rattray & V. Versace, 2012. Are we predicting the actual or apparent distribution of temperate marine fishes? *PLoS ONE* 7(4): e34558. doi:10.1371/journal.pone.0034558.
- Moore, C., E. Harvey & K. Niel, 2009. Spatial prediction of demersal fish distributions: enhancing our understanding of species–environment relationships. *ICES Journal of Marine Science*, 66:2068–2075.
- Morato, T., M. Machete, A. Kitchingman, F. Tempera, S. Lai, G. Menezes, T.J. Pitcher & R.S. Santos, 2008. Abundance and distribution of seamounts in the Azores. *Marine Ecology Progress Series*, 357:17–21.
- Morris, L. & D. Ball, 2006. Habitat suitability modeling of economically important fish species with commercial fisheries data. *ICES Journal of Marine Science*, 63: 1590–1603.
- O'Brien & Rago, 1996. An Application of the Generalized Additive Model to Groundfish Survey Data with Atlantic Cod off the Northeast Coast of the United States as an Example. *NAFO Science Council Studies*, 28:79-95.
- Olden, J. D. & D. A. Jackson, 2002. A comparison of statistical approaches for modeling fish species distributions. *Freshwater Biology*, 47: 1976–1995.
- Perkins, N.J. & E.F. Schisterman, 2006. The inconsistency of "optimal" cut-points obtained using two criteria based on the Receiver Operating Characteristic Curve. *American Journal of Epidemiology* 163: 670–675.
- Perry, A. L., P. J. Low, J. R. Ellis & J. D. Reynolds, 2005. Climate change and distribution shifts in Marine Fishes. *Science*, 308 (5730): 1912–1915.
- Peterson, A. T., M. Papes & M. Eaton, 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30: 550–560.
- Phillips, S. J., M. Dudik & R.E. Schapire, 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004.
- Phillips, S. J., R. P. Anderson & R. E. Schapire, 2006. Maximum entropy modeling of species geographic distribution. *Ecological Modelling*, 190: 231–259.
- Phillips, S. J. & M. Dudik, 2008. Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.
- Pierce, J. L. & M. S. Boyce, 2006. Modelling distribution and abundance with presence-only Data. *Journal of Applied Ecology*, 43: 405–412.

- Pitcher, T.J., T. Morato, P. J. B. Hart, M. R. Clark, N. Haggan & R. S. Santos, 2007. Seamounts: ecology, fisheries and conservation. *Fish and Aquatic Resources Series 12*. Blackwell, Oxford, pp 1–527.
- Pittmann, S. J., J. D. Christensen, C. Caldow, C. Menza & M.E. Monaco, 2006. Predictive mapping of fish species richness across shallow–water seascapes in the Caribbean. *Ecological Modelling*, 204: 9–21.
- Pittmann, S. J. & K. A. Brown, 2011. Multi–Scale approach for predicting fish species distributions across coral reef seascapes. *PLoS ONE*, 6(5): e20583. Doi:10.1371/journal.pone.0020583.
- R Development Core Team , 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL <http://www.R-project.org/>.
- Ready, J., K. Kaschner, A. B. South, P. D. Eastwood, T. Rees, J. Rius, E. Agbayani, S. Kullander & R. Froese, 2010. Predicting the distribution of marine organisms at the global scale. *Ecological Modelling*, 221: 467–478.
- Ridenour, T.A. & A.C. Heath, 1999. A note on issues in meta–analysis for behavioural genetic studies using categorical phenotypes. *Behavioural Genetics*, 29: 155–162.
- Roberts, J.J., B.D. Best, D.C. Dunn, E.A. Treml & P.N. Halpin, 2010. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software*, 25:1197–1207.
- Skov, H., E. Humphreys, S. Garthe, K. Geitner, D. Grémillet, K. C. Hamer, J. Hennenke, H. Parner & S. Wanless, 2008. Application of habitat suitability modelling to tracking data of marine animals as a means of analyzing their feeding habitats. *Ecological Modelling* 212: 504–512.
- Su N-J., S-Z. Yeh, C-L. Sun, A.E. Punt, Y. Chenc & S-P. Wang, 2008. Standardizing catch and effort data of the Taiwanese distant–water longline fishery in the western and central Pacific Ocean for bigeye tuna, *Thunnus obesus*. *Fisheries Research* 90: 235–246
- Sundermeyer, M. A., B. J. Rothschild & A. R. Robinson, 2006. Assessment of environmental correlates with the distribution of fish stocks using a spatially explicit model. *Ecological Modelling*, 197:116–132.
- Tsoar, A., O. Allouche, O. Steinitz, D. Rotem & R. Kadmon, 2007. A comparative evaluation of presence only methods for modelling species distribution. *Diversity and Distribution*, 13: 397–405.

- Valavanis, V. D., G. J. Pierce, A. F. Zuur, A. Palialexis, A. Saveliev, I. Katara & J. Wang, 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia* 612:5–20.
- Venables, W.N. & C.M. Dichmont, 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries Research* 70:319–337.
- Young, M., P. Iampietro, R. Kvitek & C. Garza, 2010. Multivariate bathymetry–derived generalized linear model accurately predicts rockfish distribution on Cordell Bank, California, USA. *Marine Ecology Progress Series* 415:247–261. doi:10.3354/meps08760