Predictive habitat modelling of reef fishes with contrasting trophic ecologies

Mara Schmiing*, Pedro Afonso, Fernando Tempera, Ricardo S. Santos

Centre of IMAR/Department of Oceanography and Fisheries of the University of the Azores and LARSyS — Associated Laboratory, 9901-862 Horta (Azores), Portugal

ABSTRACT: The success of marine spatial management and, in particular, the zonation of marine protected areas (MPAs), largely depends on the good understanding of species' distribution and habitat preferences. Yet, detailed knowledge of fish abundance is often reduced to a few sampled locations and a reliable prediction of this information across broader geographical areas is of major relevance. Generalised additive models (GAMs) were used to describe species–environment relationships and identify environmental parameters that determine the abundance or presence–absence of 11 reef fishes with contrasting life histories in shallow habitats of the Azores islands, Northeast Atlantic. Predictive models were mapped and visualised in a geographic information system (GIS) and areas with potential single or multi-species habitat hotspots were identified. Schooling, pelagic species typically required presence–absence models, whereas abundance models performed well for benthic species. Depth and distance to sediment significantly described the distribution for nearly all species, whereas the influence of exposure to swell or currents and slope of the seafloor depended on their trophic ecology. Potential presence of single species was widespread across the study area but much reduced for multiple species. There were no habitats shared by high abundances of all species in a given trophic group, and areas shared by minimal abundances were smaller than expected. Potential habitat hotspots should be considered as priority sites for conservation, but were only partially included in the existing MPA network. These findings highlight the potential of this methodology to support scientifically sound conservation planning, including but not restricted to fragmented and constrained habitats, such as those of oceanic archipelagos.

KEY WORDS: Spatial predictive modelling · Habitat mapping · Generalised additive model · GAM · Multi-species · Abundance · GIS · Azores

INTRODUCTION

The characterisation of species distribution patterns and the understanding of processes controlling these patterns reach far back into the history of biological research (reviewed in Guisan & Thuiller 2005). Recent technological advances, namely the progress in remote-sensing techniques, computing power and dissemination of information (e.g. Mumby et al. 2004, Pittman et al. 2009), together with methodological advances, including increasing capabilities of geographic information systems (GIS), a growing number of modelling tools and improved algorithms (e.g. Wood 2004, Guisan & Thuiller 2005, Swenson 2008), has facilitated the use of species distribution models. The importance of predictive habitat modelling in particular has increased, and there are a variety of techniques to model species–environment relationships and to project models into non-surveyed areas (Guisan &

Zimmermann 2000, Huettmann & Diamond 2001, Hirzel & Le Lay 2008).

Regression techniques such as parametric generalised linear models (GLMs) that relate response variables, also not normally distributed, to a linear combination of predictor variables via a link function (McCullagh & Nelder 1989, Hastie & Tibshirani 1990) are well established in marine science (e.g. Friedlander et al. 2003, Beger & Possingham 2008, Lauria et al. 2011). However, this is often not sufficient to fit complex non-linear relationships that typically occur in ecological data. Generalised additive models (GAMs) are a more flexible expansion of GLMs and use non-parametric smoothers to model species–environment relationships (e.g. Hastie & Tibshirani 1986, Murase et al. 2009). In addition, by using a suitable link function, they avoid, for instance, the need to transform count data (O'Hara & Kotze 2010). Selecting the best available model includes a thorough data exploration, the definition of an appropriate model, systematic validation and, if necessary, re-fitting of the model (e.g. Guisan & Zimmermann 2000, Zuur et al. 2009, 2010). Disregarding underlying statistical assumptions may lead to biased parameter estimates and, consequently, to incorrect ecological conclusions (e.g. Bolker et al. 2009, Zuur et al. 2010).

Statistical modelling of species' habitats has supported conservation and spatial planning and, in particular, assisted in the design of marine protected areas (MPAs). For instance, parameters that influence the functioning of MPAs are identified and assessed (e.g. Friedlander et al. 2003, Claudet et al. 2010, Vandeperre et al. 2011) and can be integrated into future MPA designs. Specifically, predictive modelling is a powerful tool because results can be spatially visualised, for instance in a GIS, allowing establishment of habitat suitability maps and identification of essential fish habitats (e.g. Pittman et al. 2007a, Bellido et al. 2008, Lauria et al. 2011). Species maps improve the use of conservation-planning tools that explore potential marine conservation scenarios and support evaluation of existing MPAs, eventual adaptation or proposal of new ones (e.g. Cañadas et al. 2005, Leathwick et al. 2008). Such illustrative and clear representations of suitability maps and conservation scenarios increase the capabilities of stakeholders and decision makers to perceive distinct spatial planning alternatives and support decision-making processes.

MPAs have become an important tool for conserving biodiversity and promoting the ecosystem-based management of sustainable fisheries (Vandeperre et al. 2008). They may represent refuge and recovery areas for depleted fish stocks where classic fisheries management methods are insufficient (Russ 2002, Higgins et al. 2008, Vandeperre et al. 2011). As MPAs may not benefit all target species equally, mainly as a result of their different habitat requirements, a multi-species approach to the optimal MPA design is strongly advisable (e.g. Possingham et al. 2000, Claudet et al. 2010).

Most studies that have used predictive modelling to characterise reef fish assemblages in support of marine management were conducted in coastal environments on continental shelves (e.g. Cañadas et al. 2005, Bellido et al. 2008, Mellin et al. 2010, Arias-González et al. 2012). In contrast, few studies consider oceanic archipelagos, typically tropical islands (e.g. Pittman et al. 2007a, Knudby et al. 2011, Pittman & Brown 2011). The Azores are a temperate oceanic, volcanic archipelago and as such have a rather unique combination of marine conditions because (1) shallow-water habitats are limited, (2) permanent natural forces are present, e.g. locally strong tidal currents and swell, and (3) island habitats (and populations) are isolated from other areas (Santos et al. 1995, Tempera 2008). Human impacts are mainly fisheries, aggregate extraction and coastal development in the Azores. Yet shallow coastal areas are essential for certain life stages of many fish species (e.g. Porteiro et al. 1996, Nash & Santos 1998, Afonso et al. 2008a) but very little is known about their habitat use (e.g. Santos & Nash 1995, Afonso et al. 2008b,c, Fontes et al. 2009). An improved understanding of coastal assemblages is of major relevance to assessing the performance of existing MPAs in the archipelago, for example by analysing if and how fish assemblages respond to different protection levels or the time period of protection (P. Afonso et al. unpubl.).

The objectives of this study were to (1) assess the spatial distribution of the most typical and most abundant coastal reef fishes, (2) understand which environmental parameters influence their abundance and presence in an oceanic archipelago, (3) identify (dis)similarities in these relationships for different trophic guilds, (4) assess the relative performance of GAMs for this purpose, (5) establish predictive maps of the potential habitat use of each species, and (6) identify areas of shared high importance for multiple species that can be potential priority areas for conservation. This study also emphasises the thorough application of GAMs and, as such, updates results from Schmiing et al. (2009).

MATERIALS AND METHODS

Study area

The Azores islands, the north-westernmost archipelago of the Macaronesia region (also including Madeira, Savage Islands, Canary Islands, Cape Verde), spread across 600 km in the Northeast Atlantic (Fig. 1). Coastal habitats, from the surface down to 40 m depth, of Faial and (western) Pico Island, located in the central group, were studied. The 2 islands are separated by a relatively narrow (6 km) and shallow (maximum depth of 190 m) passage, the 'Faial-Pico channel', that is unique in the Azores, where islands are normally separated by deep waters. The study area comprises a variety of coastal habitats, including bedrock reefs, boulder fields, cliffs, submerged cones and enclosed sandy bays, that are subject to a variety of hydrodynamic conditions. More than half of the study area (59%, 34 km^2) is composed of rocky substrate. Environmental characteristics are described in detail by Santos et al. (1995), Tempera (2008) and Tempera et al. (2012). Every island of the archipelago has designated 'Island Natural Parks' that include terrestrial and marine sites (marine park, MP) and comply with the classification of the International Union for Conservation of Nature (IUCN). Five marine sites of the Faial-MP and Pico-MP are (partly) inside the study area.

Fish counts

Fish assemblages were surveyed using underwater visual censuses (UVC) between June 1997 and October 2004 from the surface down to 40 m. Following standard belt transect surveys (Brock 1954), SCUBA divers sampled a total of 462 transects with 50 m length and 5 m width. All mobile fish were identified to the lowest possible taxon and counted, classifying individuals into size classes (juvenile, small, medium, large, very large; adapted from Bertoncini et al. 2010), according to species-specific size ranges for the Azores (Morato et al. 2001a). Sampling sites (Fig. 1) were chosen based on the general representativeness of the habitat. Transects within sites were selected by stratified random sampling: 1 to 7 transects were done in each dive without replicating the depth strata (every 10 m) or overall bottom type or crossing transects during a single dive or day (Zuur et al. 2012). Transects were georeferenced according to 'transit marks' and site proxies as recorded in the sampling protocol (Holt & Sanderson 2001) and projected in a GIS (ArcGIS 9.3 ©ESRI).

Only transects conducted in months of warmer sea temperature (June to November) were used for modelling to account for differences in sampling effort. Species used for modelling were selected based on the following criteria: species had to (1) be among the most abundant and frequent in the study area and rocky reefs from other temperate areas (e.g. Patzner & Santos 1993, Pais et al. 2007, De Raedemaecker et al. 2010), (2) have contrasting life histories (Harmelin-Vivien et al. 2001, Figueiredo et al. 2005), (3) be commercially important (7 out of 11 species; Morato et al. 2001b), (4) have ecological information available for the Azores (e.g. Afonso 2002, Afonso et al. 2008a,b,c, Fontes et al. 2009), and (5) be adequately counted using belt transect surveys (e.g. Tempera et al. 2001, Pais et al. 2007, Bertoncini et al. 2010). The final list of selected species included 5 carnivores (rainbow wrasse *Coris julis*, ballan wrasse *Labrus bergylta*, blacktail comber *Serranus atricauda*, yellowmouth barracuda *Sphyraena viridensis*, emerald wrasse *Symphodus caeruleus*), 2 herbivores (salema *Sarpa salpa*, parrotfish *Sparisoma cretense*), 2 omnivores (blue damselfish *Abudefduf luridus*, white seabream *Diplodus sargus*) and
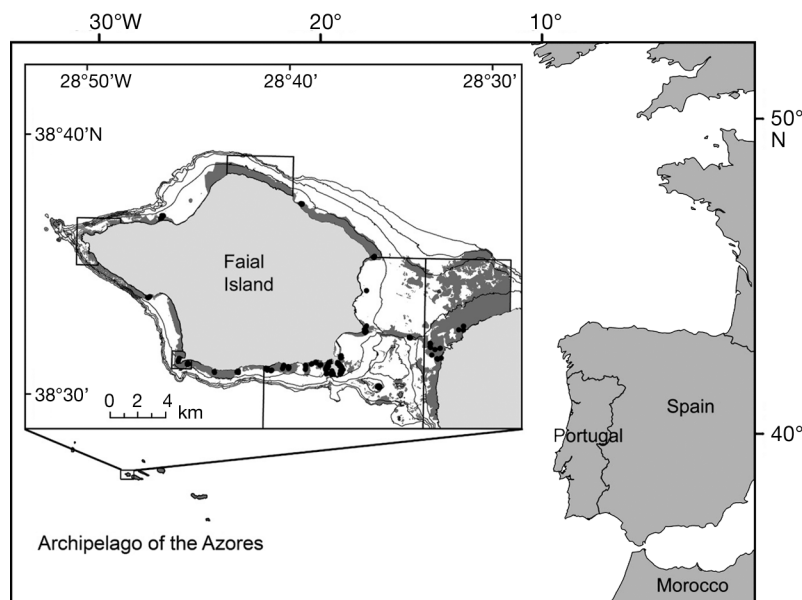

Fig. 1. Study area, showing sites of 462 surveyed transects (dots), 50 m isobaths down to 200 m and existing marine protected areas (boxes). Dark grey shaded areas represent hard-bottom mapped down to 200 m

2 planktivores (blue jack mackerel *Trachurus picturatus*, *Pagellus* spp.) (Table 1). Juveniles of these species are mostly cryptic and not efficiently counted during UVCs and were thus excluded from the analysis. Exceptions were the young-of-the-year (YOY, juveniles and small size class individuals) of *T. picturatus* and *Pagellus* spp. (including undistinguishable YOY of blackspot seabream *P. bogaraveo* and axillary seabream *P. acarne*). These species can be surveyed with UVCs because they aggregate in pelagic schools in coastal habitats of the Azores, whereas larger individuals are typical found at seamounts and in deeper waters (Menezes et al. 2006).

Environmental predictor variables

Six environmental predictor variables were available as GIS raster layers with a 5 × 5 m grid resolution: bathymetry, slope, bottom type (all 3 from multibeam surveys), exposure to currents (upscaled, from an oceanographic model), exposure to swell (upscaled, GIS-based fetch analysis), and surface chlorophyll *a* levels (upscaled, from satellite imagery) (from Tempera 2008, Tempera et al. 2012, supplementary material and a detailed description of the methods used to obtain these covariates is available online). Additional rasters (5 m cell size) were calculated in ArcGIS, namely habitat complexity parameters (maximum rate of maximum slope change, 'slope of the slope': Pittman et al. 2009; ruggedness: Sappington et al. 2007), distance to the coast, distance to the nearest hard bottom and distance to the nearest soft bottom (Table 2). Transects were inter-

sected with this suite of rasters in ArcGIS to calculate mean values of environmental predictor variables per transect, with the exception of the bathymetry grid. Given the general importance of bathymetry as predictor and the higher accuracy of *in situ* depth measurements, mean transect depth as recorded *in situ* during sampling was used as bathymetry input in the statistical models. The predictor variables distance to coast/rock/sediment were log transformed (+1 for covariates including zero values), and a square-root transformation was applied to ruggedness and exposure to currents.

Species distribution models

All statistical analyses were performed with the software R (version 2.9.1; R Development Core Team 2009). Accurate data exploration following Zuur et al. (2007, 2010) was applied to the final data set which contained a total of 354 transects. Homogeneity, potential outliers and amount of zeros were analysed with Cleveland dotplots (Cleveland 1993), boxplots and frequency plots. Spatial independency was evaluated with variograms but only for shorter distances (<1000 m) to avoid possible land intersections. Variance Inflation Factor (VIF) analysis and Spearman's rank correlations were used to test for (multi)-collinearity. Covariates with a VIF higher than 3 were removed stepwise and the analysis repeated until all values were below this cut-off level. A pairwise correlation of $\rho < -0.5$ and > 0.5 was considered high for this study, and one of the correlated covariates was removed.

Table 1. Summary of the studied reef species, including their trophic ecology (TE), observed mean number per transect (mean ± SE), highest observed count (excluding outliers; Max.), and percentage of zero counts (Zero). Model type (Model) and explained deviance (ED) of species distribution models are given. (N)B GAM: generalised additive model with a (negative) binomial distribution

Species	TE	Mean (±SE)	Max.	Zero (%)	Model	ED (%)
Coris julis	Carnivore	25.3 ± 1.7	207	19	NB GAM	67.6
Labrus bergylta	Carnivore	0.6 ± 0.1	11	69	NB GAM	43.2
Serranus atricauda	Carnivore	3.2 ± 0.2	18	25	NB GAM	45.0
Sphyraena viridensis[a]	Carnivore	3.2 ± 0.9	180	81	B GAM	43.3
Symphodus caeruleus[a]	Carnivore	1.0 ± 0.2	50	71	B GAM	34.3
Sarpa salpa	Herbivore	21.6 ± 2.4	200	38	NB GAM	43.7
Sparisoma cretense	Herbivore	4.9 ± 0.6	83	47	NB GAM	50.5
Abudefduf luridus	Omnivore	3.3 ± 0.3	31	42	NB GAM	55.6
Diplodus sargus	Omnivore	35.1 ± 4.1	500	16	NB GAM	45.8
Pagellus spp.[a, b]	Planktivore	6.0 ± 2.5	1000	82	B GAM	21.1
Trachurus picturatus[a, b]	Planktivore	377.8 ± 90.2	20000	70	B GAM	18.7

[a]Presence–absence data
[b]Juveniles and smallest size class

Table 2. Summary of environmental predictors in this study. All variables are continuous, except for bottom type

Predictor	Description	Unit	Range
Depth	Water depth	Metres	2.4–38.5
Slope[a]	Steepness of the terrain	Degrees	0.6–43.5
Swell[a]	Relative exposure to swell and waves	Percentage of maximum	0.3–72.0
Chlorophyll[a]	Average surface chlorophyll *a* concentration	Milligram per m^3	0.52–1.12
Distance to coast	Distance of transect to shore	Metres	4.0–2999.5
Distance to rock	Distance of transect to nearest hard bottom	Metres	0–138.0
Distance to sediment	Distance of transect to nearest soft bottom	Metres	0–731.8
Current[a]	Relative exposure to currents	Percentage of maximum	0–84.6
Ruggedness	Terrain ruggedness as the variation in 3D orientation of grid cells within a neighbourhood	Dimensionless; 0 (flat) to 1 (most rugged)	0–0.12
Bottom type[a]	Classification into hard and soft bottom	Dimensionless; rock = 0, sediment = 1	0, 1
Slope of slope	Maximum rate of maximum slope change	Degrees of degrees	0.8–69.2

[a]after Tempera (2008)

Fish counts were overdispersed and, as such, modelled with a negative binomial error distribution (with log link function). GAMs were applied to model the apparent non-linear relations using the mgcv package (version 1.5-5; Wood 2004, 2006). Smoothness selection of thin plate regression splines was done automatically during the model fitting process while searching for the dispersion parameter (θ) that minimised the Un-Biased Risk Estimator/Aikake Information Criterion (AIC) of the model (Wood 2004). The general formula of the full model was:

$$Y_i = \alpha + f_1(X_1) + f_2(X_2) + f_3(X_3) + \dots + f_n(X_n) + \varepsilon_i$$

with Y being the response variable, α the intercept, $f_{1-n}()$ the smoothing function for predictor variable X_{1-n}, and ε the unexplained variation for sample i.

Abundances of *Pagellus* spp., *Sphyraena viridensis*, *Symphodus caeruleus* and *Trachurus picturatus* had the highest percentage of zeros and were converted to presence–absence because initial count models exhibited residual patterns and could not be improved. Instead, a binomial error distribution with a cloglog link function was applied, except for the model for *Sphyraena viridensis* that used a logit link instead. Starting with the full model, stepwise backwards selection was used for model selection by AIC minimisation. This criterion might favour more complex models (Murase et al. 2009); consequently, hypothesis testing was used in the last step to avoid overfitting. Smoothers that continuously included zero in the confidence interval were removed (Wood & Augustin 2002). Model selection was confirmed by comparing the adjusted R^2 of nested models. The relative proportion of the deviance explained by each predictor in the final model was estimated.

Model validation

Model residuals were plotted against (1) each predictor variable (included in the model, excluded during model selection or data exploration), (2) the spatial coordinates, and (3) the temporal covariates year and month to verify the absence of residual patterns, heterogeneity, outliers, spatial or temporal dependence. Leverage was investigated graphically. Possible influential observations were removed and the model process repeated from the beginning. The dispersion parameter was estimated to confirm that overdispersion was no longer present. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) were calculated to test the accuracy of binomial models (package ROCR; Sing et al. 2005). 'Winter' data (December to May) from the same years and study area were used to examine model generality after applying a Mann-Whitney test to test for differences between winter and summer data. Predictions for this validation data set were compared to actual counts with the Spearman's rank correlation ρ and Wilcoxon signed-rank test (null hypothesis: no difference between mean ranks). The percentage of correctly predicted presences (greater than or equal to 1 fish for count models) and absences was estimated.

Spatial prediction of species' distributions

Abundances or presence of species with no residual patterns in the final model were predicted in R on a 5 m grid over the entire study area down to the 40 m isobath. In a few locations, environmental

values eventually exceeded the sampled range and predictions were extrapolated, but never for all predictor variables simultaneously. Accuracy was assessed by comparing standard errors of extrapolations and interpolations. Relative abundances (the ratio of predicted to maximum abundance) were calculated for count models to better compare species with contrasting schooling behaviour. If the modelled maximum exceeded the observed maximum (excluding outliers; Table 1), the latter was taken as a reference value to mitigate possible over-estimation. Results were imported and visualised in ArcGIS. The percentage of available rocky habitat with potential presence of single species (scoring at least a 0.1 probability of occurrence on a scale from 0 to 1 for binomial models) was estimated. This was repeated for rocky habitats that were potentially suitable for different levels of relative abundance of single species ('single-species hotspots'), setting predictive abundances with at least 10% (Max_{10}), 25% (Max_{25}) and 50% (Max_{50}) of the maximum as target levels. In a next step, maps were overlaid and areas with overlapping potential presence (Max_{10}, Max_{25}, Max_{50}; 'multi-species hotspots') of all carnivores and omnivores identified. This was not necessary for herbivores and planktivores because predictions were not done for multiple species. The location of multi-species hotspots in relation to the existing coastal MPA network was analysed.

RESULTS

Statistical assumptions and performance of GAMs

The initial set of 11 environmental predictors was reduced to 5 to avoid including correlated variables in the models. The retained predictors were depth, slope, distance to sediment (log-transformed), exposure to current (square root-transformed) and exposure to swell. It was possible to establish count models for species with up to 69% of zeros (Table 1). Final models explained 34.3 to 67.6% of the deviance, except for planktivore models that had the lowest values (≤21.1%; Table 1). No outliers or overdispersion were present and variograms did not indicate spatial correlation. Further residual plots confirmed homogeneity and independence, except for the models for *Pagellus* spp., *Sparisoma cretense* and *Sphy-*

raena viridensis which exhibited spatial clusters. A significant difference between group medians of the model fitting and validation data set was only detected for *Coris julis*. Model validation results, which generally performed well, are summarised in Table A1 in the Appendix.

Importance of environmental predictors

Depth, exposure to swell and/or current, and distance to sediment were significant predictors for 90 to 100% of the species. Distance to sediment contributed most to the total explained deviance (mean relative explained deviance: 48.5 ± 19.3%), followed by depth (24.4 ± 14.0%), exposure to current (12.7 ± 18.4%) and swell (10.7 ± 10.3%), and slope (3.6 ± 5.5%) (Fig. 2). The distribution of carnivores was described by 4 to all 5 predictor variables, whereas herbivores and omnivores responded to 3 predictor variables and planktivores to even less (2) (Figs. 2 to 5).

(Dis)similarities of distribution models between trophic guilds

Carnivores showed a predicted preference for deeper areas further away from sediment, except for *Symphodus caeruleus*. A positive response to steeper slopes was only recorded for *Sphyraena viridensis*, whereas all labrids had decreasing predictive abundances (or presence) with increasing steepness (Fig. 3). Omnivores and herbivores had similar pat-
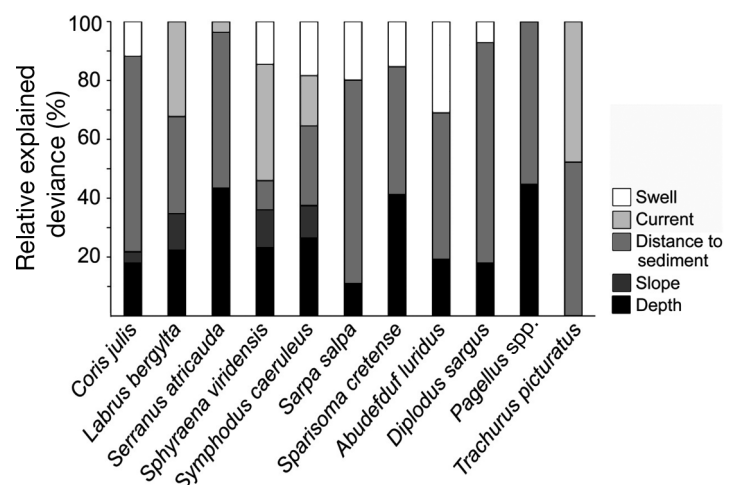

Fig. 2. Relative contribution (%) of each predictor variable to the explained deviance of 11 reef species models
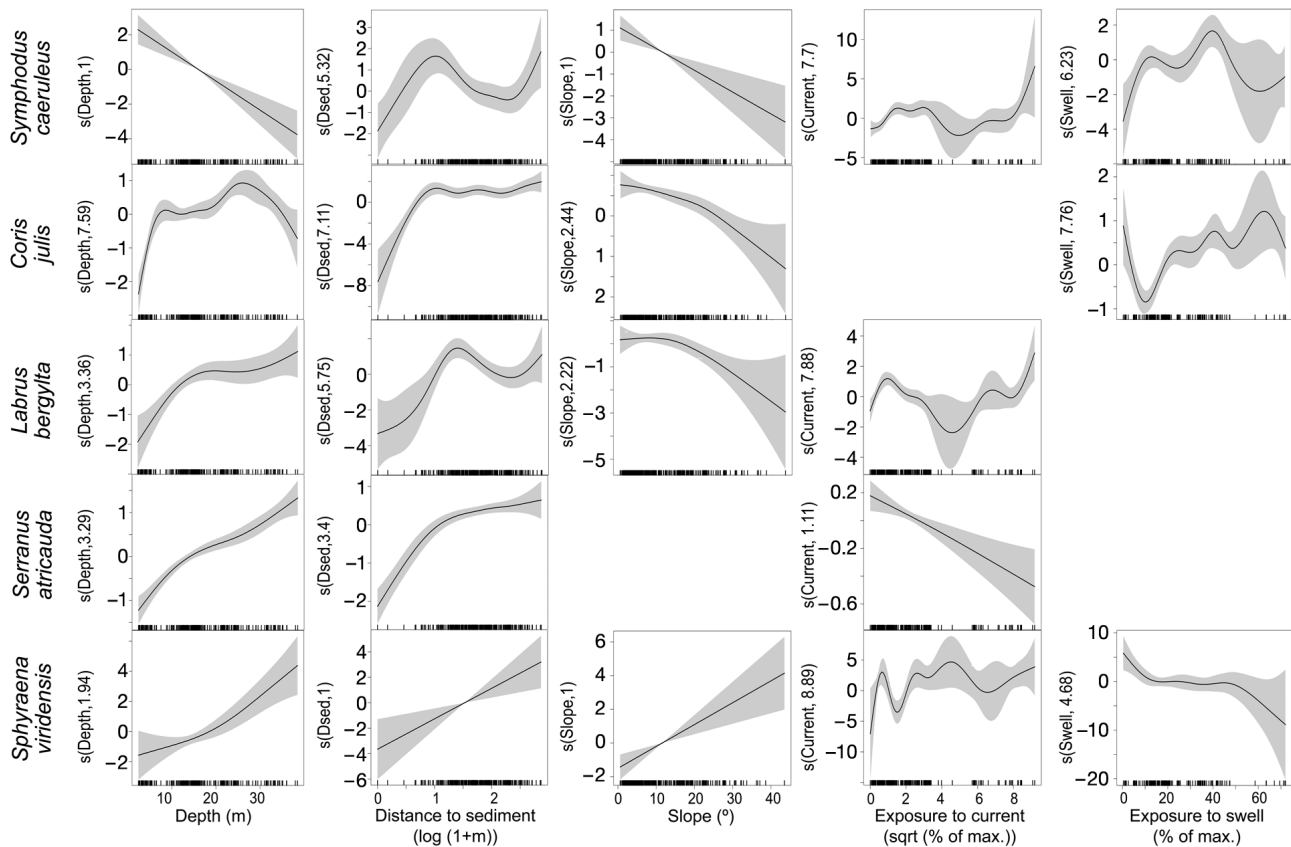
Fig. 3. Smoother estimates (solid line) for the environmental predictors depth, distance to sediment (Dsed), slope, exposure to current and swell as obtained by generalised additive models (GAMs) for carnivorous reef fishes. The approximate 95% confidence envelopes are indicated (grey shading). Tick marks on the *x*-axis are sampled data points. sqrt: square root

terns for the resulting smoothers, with the exception of predicted abundance maxima at different depths, which were also shallower in comparison to carnivores (Fig. 4). The response to distance to sediment was strongest at the intercept between bottom types (more than 10 m distance to sediment) and predictive abundances decreased in areas with highest exposure to swell. Planktivores had a predicted maximum of occurrence in the transition zone between substrates. *Pagellus* spp. was also significantly influenced by depth, whereas exposure to current influenced the occurrence of *Trachurus picturatus* (Fig. 5).

Predictive habitat maps and identification of species hotspots

Table 3 summarises calculations made with predictive maps (Figs. 6 & 7) and reports the percentage of available reef habitat that was potentially occupied by single species and different trophic groups. Individual species were predicted to occur in 21 to 98%

of the available rocky habitat. That area was much reduced for simultaneously predicted presence of herbivores or omnivores (62%) and even more for carnivores (12%). Only a minor percentage of the available reef was identified as a single-species hotspot with Max_{25} or Max_{50} as target. The herbivorous *Sarpa salpa* and the carnivorous *Coris julis* were omnipresent and the only species with large hotspots with more than 75% predicted relative abundance (Fig. 6). The centre of these hotspots had higher prediction standard errors due to extrapolation of exposure to swell (as for *Sarpa salpa*) or distance to sediment (as for *C. julis*) values. Standard errors were also higher for extrapolation of exposure to swell in predictions of *Symphodus caeruleus.* But maps did not indicate sharp contrasts between such areas and surrounding ones, instead similar distribution patterns were found in comparable locations.

Multi-species hotspots were identified by overlaying predictive abundance maps of all carnivores or omnivores. They were reduced to 12–17% of the available reef habitat and only identified for minimal abundance levels (Max_{10}), mainly at the south coast (and
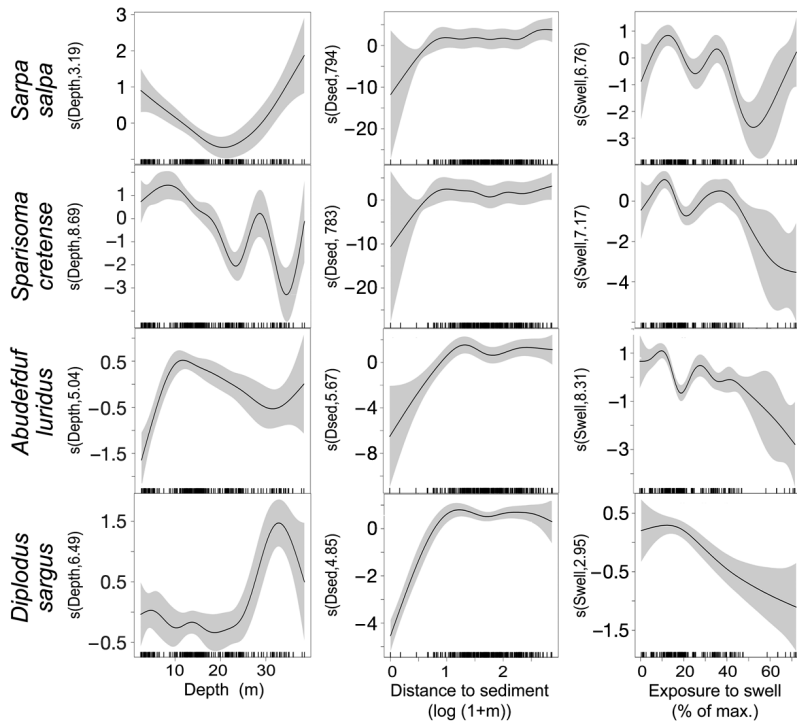
Fig. 4. Smoother estimates (solid line) as obtained by generalised additive models (GAMs) for herbivorous (*Sarpa salpa*, *Sparisoma cretense*) and omnivorous (*Abudefduf luridus*, *Diplodus sargus*) reef fishes. Other details as in Fig. 3
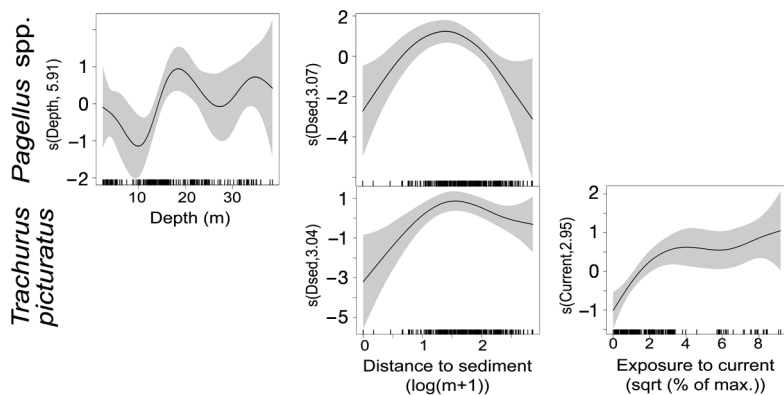

Fig. 5. Smoother estimates (solid line) as obtained by generalised additive models (GAMs) for planktivorous reef fishes. Other details as in Fig. 3

north coast for carnivores) of Faial Island (Fig. 8). At least half of these multi-species hotspots (50% for carnivores, 83% for omnivores) were located in the existing MPAs where they represented 13 to 15% of the total protected reef habitat down to 40 m.

Predictive maps for *Pagellus* spp., *Sparisoma cretense* and *Sphyraena viridensis* were not included in the analysis to avoid any inferential or misleading conclusions, although models gave plausible results

that are supported by other studies (de Girolamo et al. 1999, Barreiros et al. 2002, Afonso et al. 2008b).

DISCUSSION

This study demonstrates the adequacy of GAMs to model and predict the spatial distribution of 11 typical rocky reef fishes in the Azores. Species from the same trophic guild respond similarly to a maximum of 5 environmental predictors, with depth and distance to sediment influencing most distribution patterns. It also demonstrates how predictive models permit the mapping of species' distributions over an island scale and how hotspots with distinct abundance-targets are identified and can be considered in MPA zoning schemes.

GAM performance

Overall, GAMs performed well in modelling the species–environment relationships. However, it was not possible to find a common model type for all reef fishes. Instead, the best method reflected their habitat and schooling behaviour. Pelagic, non-territorial, (facultative) schooling species with a large yearly displacement (Claudet et al. 2010) were sighted less and typically only their presence–absence was modelled, whereas count models worked well for most benthic, solitary species. Model generality and accuracy seemed suitable, considering that they performed well in the validation, even with data from another season. However, optimally, the validation data should also be spatially independent (Araújo & Guisan 2006), but this is seldom the case, as in the present study. As GAM flexibility may lead to data overfitting, careful data exploration, model formulation, model selection and validation were applied to avoid this.

Despite this overall positive performance, 3 models reflected spatial residual patterns, indicating that an important aspect for explaining these species' distribution is missing. Results could not be improved by applying mixed models, including a spatial correla-

tion structure, or zero inflated models (see also Zuur et al. 2012 for discussion of *Sparisoma cretense*). We concluded that, rather than model misspecification, 1 (or more) missing influential predictor variable has caused the observed residual patterns. Experimental GAMs confirmed that this environmental predictor is not any of the variables excluded during data exploration, nor the spatial coordinates or other available predictors, such as protection status (e.g. Goñi et al.

Table 3. Percentage of available rocky habitat with potential presence (≥1 for count, ≥0.1 for binomial models) of single species and species from the same trophic group, as well as percentage of available rocky reefs that is potentially inhabited by 3 different levels of relative abundance (Max_{50}, Max_{25}, Max_{10}). NA: a binomial model was applied and as such relative abundances are not applicable. TE: trophic ecology (C: carnivore; H: herbivore; O: omnivore; P: planktivore)

Species	TE	Presence	Max_{50}	Max_{25}	Max_{10}
C. julis	C	98	19	49	78
L. bergylta	C	21	<1	3	19
S. atricauda	C	97	13	45	84
S. caeruleus	C	65	NA	NA	NA
S. salpa	H	94	15	25	43
A. luridus	O	62	<1	3	33
D. sargus	O	98	<1	9	37
T. picturatus	P	91	NA	NA	NA
Carnivores		12	0	1	17
Omnivores		62	0	<1	12

2008, Claudet et al. 2010) or average sea surface temperature (Tempera 2008). Residual patterns also remained if interaction between environmental variables was considered.

Predictors for habitat complexity, in particular, are known to influence fish assemblages both in temperate (Charbonnel et al. 2002, De Raedemaecker et al. 2010, Monk et al. 2011) and tropical reefs (e.g. Friedlander & Parrish 1998, Nemeth & Appeldoorn 2009, Pittman et al. 2009) and, consequently, such variables were calculated in ArcGIS. Despite their apparent importance all were excluded during data exploration because of their correlation with slope.

Nevertheless, final predictor variables were highly significant and should not lose their importance even if the missing variable is identified, measured and integrated into the model.

Environmental predictors that influence the distribution of single reef species and trophic guilds

This study concurs with results from others (e.g. Friedlander et al. 2003, Beger & Possingham 2008, De Raedemaecker et al. 2010, Krajewski & Floeter 2011) that indicated oceanographic forces (swell or current), depth and distance to sediment significantly influence the distribution of many reef fishes. How-
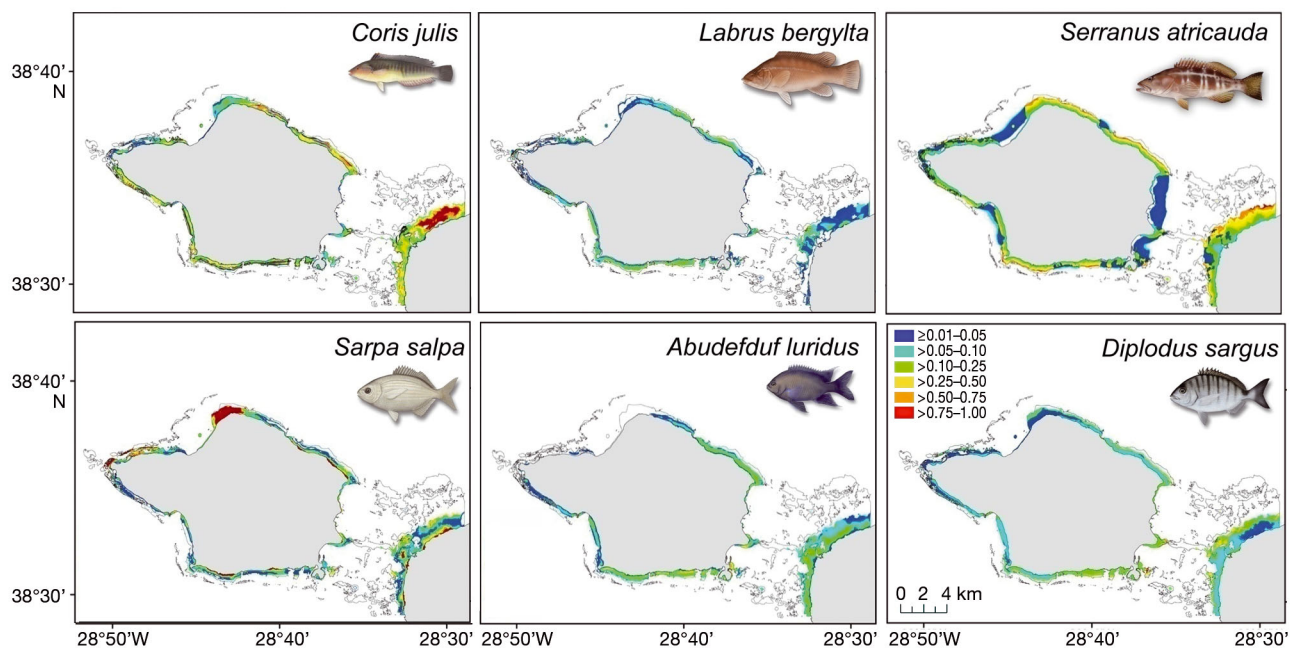

Fig. 6. Predictive maps for relative abundances of carnivorous (*Coris julis*, *Labrus bergylta*, *Serranus atricauda*), herbivorous (*Sarpa salpa*) and omnivorous (*Abudefduf luridus*, *Diplodus sargus*) reef fishes down to 40 m, including the outline of rocky substrate down to 200 m. Fish illustrations (by L. Gallagher/fishpics/ImageDOP) are not to scale
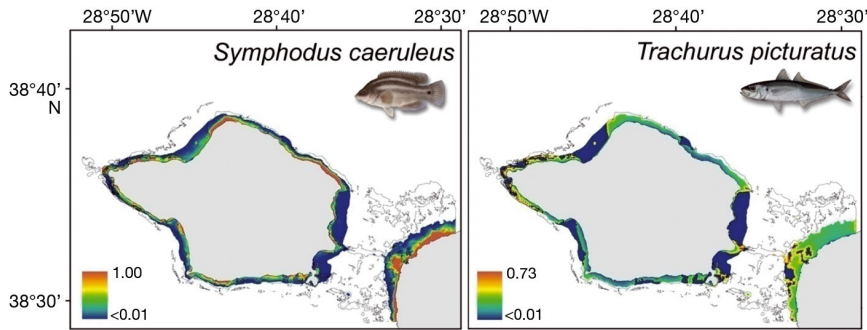
Fig. 7. Probability of presence of the carnivore *Symphodus caeruleus* and the planktivore *Trachurus picturatus* for coastal habitats down to 40 m, including the outline of rocky substrate down to 200 m. Note the different probability scales. Fish illustrations (by L. Gallagher/fishpics/Image DOP) are not to scale
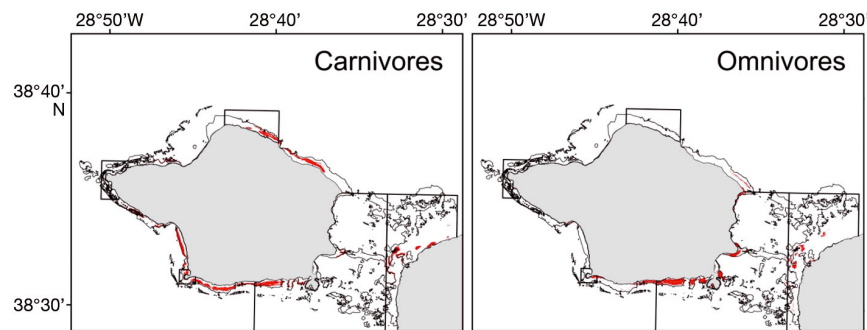

Fig. 8. Shared potential habitat (marked red) with a target of Max_{10} relative abundance of carnivores (*Coris julis*, *Labrus bergylta*, *Serranus atricauda*) and omnivores (*Abudefduf luridus*, *Diplodus sargus*) down to 40 m depth. The outline of rocky substrate down to 200 m and marine protected areas are shown

partly corroborate findings from the Mediterranean. Similar to the results of De Raedemaecker et al. (2010), no overall pattern could be distinguished for carnivores and potential presence of planktivores was least for most sheltered habitats. But on the contrary, predicted abundances of herbivores and omnivores decreased with increasing exposure.

The single predictors and their importance for individual species and trophic guilds are discussed in the following.

Substrate-related effects

The importance of seascape proxies or distance-to-domain boundaries in describing fish assemblages has been demonstrated recently (Mellin et al. 2010, Kendall et al. 2011, Pittman & Brown 2011). Similarly, distance to sediment was the most important and single common predictor. Although it is easy to measure (Friedlander & Parrish 1998), to our knowledge, it has not been used in the same way elsewhere. Studies in the Caribbean Sea used the percentage or area of sediment in proximity to coral reefs and demonstrated a correlation between these predictors and fish density, richness, or individual species, whereas correlation strength varied with the analysis radius and map resolution (Pittman et al. 2007b, Kendall et al. 2011). Friedlander & Parrish (1998) used the distance to the coral reef edge to describe fish assemblages in Hawaii and concluded that abundance, biomass and diversity are higher in close proximity to the edge. Results in the Azores show that herbivores, omnivores and carnivores have highest predicted abundances close to the sediment. Likewise, highest probabilities for the presence of planktivores (such as *Trachurus picturatus*) are at the interface between substrate types, which concurs with the observations of local fishermen (T. Morato pers. comm.). Most likely, distance to sediment has an indirect effect and the zone may be a preferred foraging ground due to local hydrodynamics and increased quantity and/or diversity in food resources. Such a complementary effect through the presence of different habitat types in close proximity is known

ever, it also shows that the species' trophic ecology shapes such relationships differently. In agreement with other studies, carnivores and planktivores respond positively to bathymetry (Friedlander & Parrish 1998, Bertoncini et al. 2010, Krajewski & Floeter 2011), whereas herbivores and omnivores are also more abundant in shallow and intermediate depths (Friedlander & Parrish 1998, Nemeth & Appeldoorn 2009). Distance to sediment influences the distribution of all trophic guilds, with increasing or highest predictive abundances (or potential presence) in proximity to soft bottom (approximately 10 m distance to sediment; Figs. 3, 4 & 5). Such maxima at habitat edges have also been reported for coral or seagrass habitats in the Caribbean (e.g. Pittman et al. 2007a, Kendall et al. 2011), Australia (e.g. Vanderklift et al. 2007, Smith et al. 2011) and the Indian Ocean (e.g. Dorenbosch et al. 2005). Slope only affects the distribution of carnivores, with the single pelagic representative being the only carnivore responding positively to it. Responses to exposure

from other studies (Pittman et al. 2007b). This significance of edge effects in the marine realm (Dorenbosch et al. 2005, Pittman et al. 2007a, Smith et al. 2011) emphasises the importance of including a continuous predictor that represents the distance to habitat edges in distribution models for reef species. Such influence would not be captured if, for instance, a categorical substrate-type predictor was used.

Substrate slope shapes hydrodynamic forces, illumination and sedimentation patterns, and as such, influences the distribution of benthic and benthopelagic assemblages (e.g. Denny et al. 2003, Letourneur et al. 2003, Milazzo et al. 2011). Its importance seems to be accentuated by the correlation with terrain complexity measures in this study, implying that a significant slope effect could also represent, for instance, a ruggedness effect. Fine-scale or *in situ* measurement of habitat complexity may even outperform slope in the study area. On the other hand, its importance might be more evident at broader spatial scales. However, it is only significant for carnivores, with relationships that corroborate other studies in the Azores, the Canaries and the Mediterranean (Barreiros et al. 2002, Letourneur et al. 2003, Milazzo et al. 2011). Pittman & Brown (2011) highlight not only the significance of such substrate-related predictors but also the importance of interactions between the geographical location (i.e. distance to sediment) and the topographic heterogeneity of seafloor structure (i.e. slope). However, interactions are not used in our study because experimental GAMs revealed that resulting 2D-smoothers were either not significant or not interpretable or meaningful and seemed to overfit the data.

Depth-related effects

Depth was the second most important predictor. The bathymetric distribution patterns corroborate previous findings for the Azores (e.g. Patzner & Santos 1993, Harmelin-Vivien et al. 2001, Bertoncini et al. 2010) and other areas (e.g. Nemeth & Appeldoorn 2009, Krajewski & Floeter 2011, Richards et al. 2012), and indicate that carnivorous species (*Coris julis*, *Labrus bergylta*, *Serranus atricauda*, *Sphyraena viridensis*) have higher predictive abundances for deeper habitats, whereas herbivorous species (*Sarpa salpa*, *Sparisoma cretense*) prefer shallow or intermediate depths. Planktivores responded positively to increasing depth in Hawaii (Friedlander & Parrish 1998), which was also observed for the planktivorous *Pagellus* spp. in the Azores. The importance of

bathymetry for this guild would probably be more pronounced (i.e. for *Trachurus picturatus*) if habitats below 40 m were included (Menezes et al. 2006).

Depth is a proxy for other environmental variables such as light penetration (Nemeth & Appeldoorn 2009), which influences the distribution and species composition of algae in the Azores (Neto 2001, Tempera 2008, Tempera et al. 2012). This, in turn, affects spatial and bathymetric patterns for herbivores (Ruitton et al. 2000) and species such as *Abudefduf luridus* and *Symphodus caeruleus* which use algae for nest building (Porteiro et al. 1996, Azevedo et al. 1999).

Bathymetric patterns are influenced by varying diet preferences of similar species or different life stages of 1 species and/or ontogenetic depth preferences (Porteiro et al. 1996, Letourneur et al. 2003, Figueiredo et al. 2005). Such niche separation has been observed for carnivorous labrids that mainly feed on small benthic crustaceans (*Symphodus caeruleus*; Azevedo et al. 1999) or sea urchins, including different sizes or species (*Coris julis*, *Labrus bergylta*; Sala 1997, Figueiredo et al. 2005).

Preference for different habitat types has also been observed for herbivores. For instance, larger individuals of *Sparisoma cretense* are typically territorial and prefer more exposed, deeper reef sections, whereas smaller individuals often form schools in sheltered, shallower habitats (de Girolamo et al. 1999, Afonso et al. 2008b). This dual social system seems to be reflected by the prediction of 2 abundance maxima for different depths and exposures to swell. Similar habitat segregation is predicted for *Sarpa salpa* and confirms findings of Harmelin-Vivien et al. (2001), who recorded lower abundances for intermediate depth strata. Other studies have shown that *Sarpa salpa* actively forages in shallow waters and shelters in deeper areas (Ruitton et al. 2000, Jadot et al. 2002), which are generally preferred by larger specimens (Afonso 2002). Given the ecological importance of this species, and that studies have so far focused on its shallower habitat use (e.g. Verlaque 1990, Ruitton et al. 2000, Jadot et al. 2002, 2006), more research is needed to understand its use of deeper areas.

Exposure-related effects

Exposure to swell or currents were significant for the distribution of all but 1 species. Similar hydrodynamic influences on reef fishes have been demonstrated in other volcanic archipelagos (e.g. Friedlan-

der et al. 2003, Beger & Possingham 2008, Krajewski & Floeter 2011, Richards et al. 2012).

Pais et al. (2007) reported significantly higher abundance of carnivorous labrids and serranids in sheltered habitats in the Mediterranean, whereas De Raedemaecker et al. (2010) did not find a significant difference for these guilds. However, those studies used different exposure measurement scales, and overall exposure is greater in the Azores. Carnivores, especially, vary in their exploitation of different exposure regimes in the archipelago, which may partly be explained by species-specific swimming performances (e.g. Fulton & Bellwood 2004, Fulton et al. 2005). Species with high aspect ratio fins have enhanced swimming abilities, swim faster and are likely to also inhabit exposed environments as observed, for instance, for different *Coris* species (Wainwright et al. 2002, Fulton & Bellwood 2004, Milazzo et al. 2011). Such conditions are typically met at the offshore reefs in the Faial-Pico channel, where schools of *C. julis* are very numerous (Afonso 2002). On the contrary, *Serranus* species swim in bursts when hunting (Wakeling & Johnston 1998, authors' pers. obs.) but otherwise are sustained swimmers, explaining their preference for sheltered habitats (Dominici-Arosemena & Wolff 2005, Pais et al. 2007, De Raedemaecker et al. 2010, present study). Multiple peaks in the smoothers of exposure variables (and others) may indicate the interaction and dominance of another variable or competition and displacement by other species over exposure levels between maxima. For instance, the dual social system of the herbivorous *Sparisoma cretense* seems to be reflected here (see previous subsection; Afonso et al. 2008b). However, in general, herbivores and omnivores have non-linear decreasing predicted abundances (presence) for higher exposure regimes, contrary to planktivores. Currents carry suspended food items and as such are particular important for YOY of species such as *Trachurus picturatus*.

Multi-species habitat mapping and potential benefit for conservation planning

Priority areas for each species were identified with predictive maps that are rather different from each other. Some illustrate a strong bathymetric dependence (*Serranus atricauda*, *Symphodus caeruleus*), while the influence of the sediment–rock interface is most evident, for instance, for *Trachurus picturatus*. Its spatial pattern also overlaps with areas of highest exposure to current, such as the channel margins and

the western shore of Faial Island (Tempera et al. 2012). Such sites should have diverse and abundant food resources (Dominici-Arosemena & Wolff 2005). When interpreting these maps, it is important to keep in mind that they are restricted by the study area limits and, for a few species, by the low explained model deviance. For instance, the highest presence-probability of *T. picturatus* is only 0.73, reflecting that its main habitat at deeper strata (Menezes et al. 2006) is not integrated in the model.

Models of all 8 species selected for spatial prediction carefully comply with statistical considerations (i.e. exclusion of outliers, overdispersion, independence, heterogeneity), assuring their applicability for habitat maps. Distribution of most species from 1 trophic guild is described by the same predictor variables with similar (mostly) non-linear patterns. Yet niche separation is reflected, for instance, by different bathymetric preferences. Based on these results, it is not possible to identify, for instance, a single species as surrogate for a trophic group, but rather all 8 species act as surrogates.

Trophic guild analysis can support studies on ecosystem structure and functioning (Cartes et al. 2002, Claudet et al. 2010), even if euryphagy and ontogenetic or seasonal diet changes in reef fishes (e.g. Verlaque 1990, Azevedo et al. 1999, Figueiredo et al. 2005) pose some limitations to this use. For instance, fishing pressure can influence trophic groups differently (Ashworth & Ormond 2005). Future studies should include a larger number of species which should allow a more refined classification, for example, into different types of carnivores (Friedlander & Parrish 1998, Ferreira et al. 2001, De Raedemaecker et al. 2010) and consider other life history traits (Dominici-Arosemena & Wolff 2005, Claudet et al. 2010). For instance, model results indicate that potential habitats of solitary species with smaller home ranges (*Abudefduf luridus*, *Labrus bergylta*, *Symphodus caeruleus*) are spatially more restricted than potential rocky habitats of e.g. schooling specimens (*Coris julis*, *Diplodus sargus*, *Sarpa salpa*; Table 3). Such a refinement should allow good identification of surrogate species.

Although individual species often had a wide potential distribution, predictive maps illustrate that areas with highest relative abundance (>50% of the maxima) and, consequently, the most favourable habitat, are rather small. Only 2 hotspots — larger areas with more than 75% relative abundance of a single species — were identified, off the north-west of Faial Island (for *Coris julis*) and eastern coast of Pico Island (for *Sarpa salpa*), but these had higher

prediction standard errors in their centres. Therefore, these single-species hotspots need to be confirmed by further field surveys.

Nevertheless, the methods and results presented allow identification of important fish habitats, alias species hotspots, that ought to be considered in MPA site selection or evaluation of existing zoning schemes. For instance, rocky reef areas with potential highest relative abundances of single species (Max_{50}) were identified in the present study. Such single-species hotspots are potential priority areas for conservation and should be considered if the goal of an MPA is the protection of a single endangered species. However, multi-species spatial management is often the aim of MPA establishment, and a multitude of habitat maps has to be considered. Overlaying predictive maps for different species reduced the size of such multi-species hotspots, which was especially pronounced for abundance maps (and not merely presence). Carnivore or omnivore multi-species hotspots with a target of Max_{10} were identified off the south and north coast of Faial. These areas were much smaller than anticipated, stressing their relevance as priority sites for conservation. A subsequent analysis focused on the percentage of the hotspots that is included in the MPA network covering the study area. About 64 % (22 km^2) of the entire available reef habitat in the study area down to 40 m is encompassed in existing MPAs. Although at least half of the multi-species hotspots (carnivores: 50 %; omnivores: 83 %) are located in these managed sites, importantly, such hotspots represent only 13 % (carnivores) to 15 % (omnivores) of the reef habitat in the MPAs. Thus, any future refinement of the current zoning should target those hotspots, namely by including them entirely. However, an *in situ* evaluation of predicted hotspots should be conducted to ensure reliability and support analyses of network representativeness. Expansion to another island will help validate the presented approach for the identification of areas of priority for conservation.

This demonstration gives only a relatively simple simulation for the application of habitat maps in the context of marine conservation. Multi-species hotspots targeted only 10 % of the maximum abundance of species from the same trophic group. Other targets, namely Max_{25} or Max_{50}, were not feasible because shared habitat was diminutive. However, one might define other criteria, for instance not selecting overlapping but rather neighbouring sites with Max_{50} as target. If habitat maps with potential presence (and not abundance levels) are considered, it is crucial to define an appropriate probability level for

presence (i.e. 0.1 or 0.5). However, such decisions depend on the goal of the MPA and are not the focus of this study. Furthermore, information such as the importance of a site for recreational or fishing purposes needs to be considered as well. Site-selection or MPA evaluation is more challenging than the simple approach presented in this study. Thus, species distribution maps can be integrated in well developed multicriteria decision-making software (e.g. Marxan; Ball et al. 2009) that support MPA design by providing protection scenarios that meet distinct conservation objectives. In conclusion, adequate predictive modelling can improve the knowledge of species' spatial habitat use that, in turn, supports scientifically sound conservation planning and site selection.

LITERATURE CITED

Afonso P (2002) Spatial patterns in the littoral fish community of the Azores. MSc thesis, University of Coimbra, Coimbra, Portugal (in Portuguese with English abstract)

Afonso P, Tempera F, Menezes G (2008a) Population structure and habitat preferences of red porgy (*Pagrus pagrus*) in the Azores, central north Atlantic. Fish Res 93: 338–345

Afonso P, Morato T, Santos RS (2008b) Spatial patterns in reproductive traits of the temperate parrotfish *Sparisoma cretense.* Fish Res 90:92–99

Afonso P, Morato T, Fontes J, Holland K, Santos RS (2008c) Reproduction and spawning habitat of white trevally, *Pseudocaranx dentex*, in the Azores, central north Atlantic. Sci Mar 92:373–381

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33:1677–1688

Arias-González J, Acosta-González G, Membrillo N, Garza-Pérez J, Castro-Pérez J (2012) Predicting spatially explicit coral reef fish abundance, richness and Shannon Weaver index from habitat characteristics. Biodivers Conserv 21:115–130

Ashworth JS, Ormond RFG (2005) Effects of fishing pressure and trophic group on abundance and spillover across boundaries of a no-take zone. Biol Conserv 121:333–344

Azevedo JMN, Cepeda HIR, Rodrigues JB (1999) Notes on the biology of *Centrolabrus caeruleus* Azevedo, 1999 (Teleostei: Labridae). Arquipel Life Mar Sci 17A:27–36

Ball IR, Possingham HP, Watts M (2009) Marxan and relatives: software for spatial conservation prioritisation. In: Moilanen A, Wilson KA, Possingham HP (eds) Spatial conservation prioritisation: quantitative methods and computational tools. Oxford University Press, Oxford, p 185–195

Barreiros JP, Santos RS, De Borba AE (2002) Food habits, schooling and predatory behaviour of the yellowmouth barracuda, *Sphyraena viridensis* (Perciformes: Sphyraenidae) in the Azores. Cybium 26:83–88

Beger M, Possingham HP (2008) Environmental factors that influence the distribution of coral reef fishes: modeling occurrence data for broad-scale conservation and management. Mar Ecol Prog Ser 361:1–13

Bellido JM, Brown AM, Valavanis VD, Giráldez A, Pierce GJ, Iglesias M, Palialexis A (2008) Identifying essential fish habitat for small pelagic species in Spanish Mediterranean waters. Hydrobiologia 612:171–184

Bertoncini AA, Machado LF, Barreiros JP, Hostim-Silva M, Verani JR (2010) Rocky reef fish community structure in two Azorean islands (Portugal) central North Atlantic. J Mar Biol Assoc UK 90:1353–1362

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135

Brock VE (1954) A preliminary report on a method of estimating reef fish populations. J Wildl Manag 18:297–308

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

Cartes JE, Abelló P, Lloris D, Carbonell A, Torres P, Maynou F, Sola LGD (2002) Feeding guilds of western Mediterranean demersal fish and crustaceans: an analysis based on a spring survey. Sci Mar 66:209–220

Charbonnel E, Serre C, Ruitton S, Harmelin JG, Jensen A (2002) Effects of increased habitat complexity on fish assemblages associated with large artificial reef units (French Mediterranean coast). ICES J Mar Sci 59: S208–S213

Claudet J, Osenberg CW, Domenici P, Badalamenti F and others (2010) Marine reserves: fish life history and ecological traits matter. Ecol Appl 20:830–839

Cleveland WS (1993) Visualizing data. Hobart Press, Summit, NJ

de Girolamo M, Scaggiante M, Rasotto MB (1999) Social organization and sexual pattern in the Mediterranean parrotfish *Sparisoma cretense* (Teleostei: Scaridae). Mar Biol 135:353–360

De Raedemaecker F, Miliou A, Perkins R (2010) Fish community structure on littoral rocky shores in the eastern Aegean Sea: effects of exposure and substratum. Estuar Coast Shelf Sci 90:35–44

Denny MW, Miller LP, Stokes MD, Hunt LJH, Helmuth BST (2003) Extreme water velocities: topographical amplification of wave-induced flow in the surf zone of rocky shores. Limnol Oceanogr 48:1–8

Dominici-Arosemena A, Wolff M (2005) Reef fish community structure in Bocas del Toro (Caribbean, Panama): gradients in habitat complexity and exposure. Caribb J Sci 41: 613–637

Dorenbosch M, Grol MGG, Nagelkerken I, van der Velde G

(2005) Distribution of coral reef fishes along a coral reef–seagrass gradient: edge effects and habitat segregation. Mar Ecol Prog Ser 299:277–288

Ferreira CEL, Gonçalves JEA, Coutinho R (2001) Community structure of fishes and habitat complexity on a tropical rocky shore. Environ Biol Fishes 61:353–369

Figueiredo M, Morato T, Barreiros JP, Afonso P, Santos RS (2005) Feeding ecology of the white seabream, *Diplodus sargus*, and the ballan wrasse, *Labrus bergylta*, in the Azores. Fish Res 75:107–119

Fontes J, Caselle JE, Afonso P, Santos RS (2009) Multi-scale recruitment patterns and effects on local population size of a temperate reef fish. J Fish Biol 75:1271–1286

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

Friedlander AM, Brown EK, Jokiel PL, Smith WR, Rodgers KS (2003) Effects of habitat, wave exposure, and marine protected area status on coral reef fish assemblages in the Hawaiian archipelago. Coral Reefs 22:291–305

Fulton CJ, Bellwood DR (2004) Wave exposure, swimming performance, and the structure of tropical and temperate reef fish assemblages. Mar Biol 144:429–437

Fulton CJ, Bellwood DR, Wainwright PC (2005) Wave energy and swimming performance shape coral reef fish assemblages. Proc R Soc Lond B Biol Sci 272:827–832

Goñi R, Adlerstein S, Alvarez-Berastegui D, Forcada A and others (2008) Spillover from six western Mediterranean marine protected areas: evidence from artisanal fisheries. Mar Ecol Prog Ser 366:159–174

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8: 993–1009

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

Harmelin-Vivien ML, Harmelin JG, Almeida AJ (2001) Structure of fish assemblages on coastal rocky shores of the Azores. Bol Mus Munic Funchal 6:127–138

Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1:297–318

Hastie T, Tibshirani RJ (1990) Generalized additive models. Chapman & Hall/CRC, London

Higgins RM, Vandeperre F, Perez-Ruzafa A, Santos RS (2008) Priorities for fisheries in marine protected area design and management: Implications for artisanal-type fisheries as found in southern Europe. J Nat Conserv 16: 222–233

Hirzel AH, Le Lay G (2008) Habitat suitability modelling and niche theory. J Appl Ecol 45:1372–1381

Holt R, Sanderson B (2001) Procedural guideline No. 6-2. Relocation of intertidal and subtidal sites. In: Davies J, Baxter J, Bradley M, Connor D and others (eds) Natura 2000 — Marine monitoring handbook. Joint Nature Conservation Committee, Peterborough, p 381–397

Huettmann F, Diamond AW (2001) Seabird colony locations and environmental determination of seabird distribution: a spatially explicit breeding seabird model for the Northwest Atlantic. Ecol Model 141:261–298

Jadot C, Ovidio M, Voss J (2002) Diel activity of *Sarpa salpa* (Sparidae) by ultrasonic telemetry in a *Posidonia oceanica* meadow of Corsica (Mediterranean Sea). Aquat Living Resour 15:343–350

Jadot C, Donnay A, Acolas ML, Cornet Y, Anras MLB (2006) Activity patterns, home-range size, and habitat utilization of *Sarpa salpa* (Teleostei: Sparidae) in the Mediterranean Sea. ICES J Mar Sci 63:128–139

Kendall MS, Miller TJ, Pittman SJ (2011) Patterns of scale-

dependency and the influence of map resolution on the seascape ecology of reef fish. Mar Ecol Prog Ser 427: 259–274

Knudby A, Roelfsema C, Lyons M, Phinn S, Jupiter S (2011) Mapping fish community variables by integrating field and satellite data, object-based image analysis and modeling in a traditional Fijian fisheries management area. Remote Sens 3:460–483

Krajewski JP, Floeter SR (2011) Reef fish community structure of the Fernando de Noronha Archipelago (Equatorial Western Atlantic): the influence of exposure and benthic composition. Environ Biol Fishes 92:25–40

Lauria V, Vaz S, Martin CS, Mackinson S, Carpentier A (2011) What influences European plaice (*Pleuronectes platessa*) distribution in the eastern English Channel? Using habitat modelling and GIS to predict habitat utilization. ICES J Mar Sci 68:1500–1510

Leathwick J, Moilanen A, Francis M, Elith J and others (2008) Novel methods for design and evaluation of marine protected areas in offshore waters. Conserv Lett 1:91–102

Letourneur Y, Ruitton S, Sartoretto S (2003) Environmental and benthic habitat factors structuring the spatial distribution of a summer infralittoral fish assemblage in the north-western Mediterranean Sea. J Mar Biol Assoc UK 83:193–204

McCullagh P, Nelder JA (1989) Generalized linear models. Chapman & Hall/CRC, London

Mellin C, Bradshaw CJA, Meekan MG, Caley MJ (2010) Environmental and spatial predictors of species richness and abundance in coral reef fishes. Glob Ecol Biogeogr 19:212–222

Menezes GM, Sigler MF, Silva HM, Pinho MR (2006) Structure and zonation of demersal fish assemblages off the Azores Archipelago (mid-Atlantic). Mar Ecol Prog Ser 324:241–260

Milazzo M, Palmeri A, Falcón JM, Badalamenti F and others (2011) Vertical distribution of two sympatric labrid fishes in the western Mediterranean and eastern Atlantic rocky subtidal: local shore topography does matter. Mar Ecol 32:521–531

Monk J, Ierodiaconou D, Bellgrove A, Harvey E, Laurenson L (2011) Remotely sensed hydroacoustics and observation data for predicting fish habitat suitability. Cont Shelf Res 31:S17–S27

Morato T, Afonso P, Lourinho P, Barreiros JP, Santos RS, Nash RDM (2001a) Length-weight relationships for 21 coastal fish species of the Azores, north-eastern Atlantic. Fish Res 50:297–302

Morato T, Guénette S, Pitcher T (2001b) Fisheries of the Azores, 1982–1999. In: Zeller D, Watson R, Pitcher T, Pauly D (eds) Fisheries impacts on north Atlantic ecosystems: catch, effort and national/regional data sets. Fisheries Centre Research Reports, University of British Columbia, 9(3):214–220

Mumby PJ, Skirving W, Strong AE, Hardy JT and others (2004) Remote sensing of coral reefs and their physical environment. Mar Pollut Bull 48:219–228

Murase H, Nagashima H, Yonezaki S, Matsukura R, Kitakado T (2009) Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. ICES J Mar Sci 66: 1417–1424

Nash RDM, Santos RS (1998) Seasonality in diel catch rate of small fishes in a shallow-water fish assemblage at Porto Pim Bay, Faial, Azores. Estuar Coast Shelf Sci 47: 319–328

Nemeth M, Appeldoorn R (2009) The distribution of herbivorous coral reef fishes within fore-reef habitats: the role of depth, light and rugosity. Caribb J Sci 45:247–253

Neto AI (2001) Macroalgal species diversity and biomass of subtidal communities of São Miguel (Azores). Helgol Mar Res 55:101–111

O'Hara RB, Kotze DJ (2010) Do not log-transform count data. Methods Ecol Evol 1:118–122

Pais A, Azzurro E, Guidetti P (2007) Spatial variability of fish fauna in sheltered and exposed shallow rocky reefs from a recently established Mediterranean Marine Protected Area. Ital J Zool 74:277–287

Patzner RA, Santos RS (1993) Ecology of rocky littoral fishes of the Azores. Cour Forschinst Senckenb 159:423–427

Pittman SJ, Brown KA (2011) Multi-scale approach for predicting fish species distributions across coral reef seascapes. PLoS ONE 6:e20583

Pittman SJ, Christensen JD, Caldow C, Menza C, Monaco ME (2007a) Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. Ecol Model 204:9–21

Pittman SJ, Caldow C, Hile SD, Monaco ME (2007b) 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 Spec Issue 53:27–38

Porteiro FM, Barreiros JP, Santos Ricardo S (1996) Wrasses (Teleostei: Labridae) of the Azores. Arquipel Life Mar Sci 14A:23–40

Possingham H, Ball I, Andelman S (2000) Mathematical methods for identifying representative reserve networks. In: Ferson S, Burgman M (eds) Quantitative methods for conservation biology. Springer, New York, NY, p 291–305

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Richards BL, Williams ID, Vetter OJ, Williams GJ (2012) Environmental factors affecting large-bodied coral reef fish assemblages in the Mariana Archipelago. PLoS ONE 7:e31374

Ruitton S, Francour P, Boudouresque CF (2000) Relationships between algae, benthic herbivorous invertebrates and fishes in rocky sublittoral communities of a temperate sea (Mediterranean). Estuar Coast Shelf Sci 50: 217–230

Russ GR (2002) Yet another review of marine reserves as reef fishery management tools. In: Sale PF (ed) Coral reef fishes: dynamics and diversity in a complex ecosystem. Academic Press, San Diego, CA, p 421–443

Sala E (1997) Fish predators and scavengers of the sea urchin *Paracentrotus lividus* in protected areas of the north-west Mediterranean Sea. Mar Biol 129:531–539

Santos RS, Nash RDM (1995) Seasonal changes in a sandy beach fish assemblage at Porto Pim, Faial, Azores. Estuar Coast Shelf Sci 41:579–591

Santos RS, Hawkins S, Monteiro LR, Alves M, Isidro EJ (1995) Marine research, resources and conservation in the Azores. Aquat Conserv 5:311–354

Sappington JM, Longshore KM, Thompson DB (2007) Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. J Wildl Manag 71:1419–1426

Schmiing M, Afonso P, Tempera F, Santos R (2009) Integrat-

ing recent and future marine technology in the design of marine protected areas—the Azores as case study. In: Proc IEEE Oceans Conf Eur 2009, Bremen, doi:10.1109/OCEANSE.2009.5278178

Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21:3940–3941

Smith TM, Hindell JS, Jenkins GP, Connolly RM, Keough MJ (2011) Edge effects in patchy seagrass landscapes: the role of predation in determining fish distribution. J Exp Mar Biol Ecol 399:8–16

Swenson NG (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. J Evol Biol 21: 421–434

Tempera F (2008) Benthic habitats of the extended Faial island shelf and their relationship to geologic, oceanographic and infralittoral biologic features. PhD dissertation, University of St. Andrews. Available from http://research-repository.st-andrews.ac.uk/handle/10023/726

Tempera F, Afonso P, Morato T, Prieto R and others (2001) Biological assemblages of the Faial-Pico Channel SACs. Arquivos do DOP, Série Relatórios Internos, no. 7/2001, Departamento de Oceanografia e Pescas da Universidade dos Açores, Horta (in Portuguese)

Tempera F, McKenzie M, Bashmachnikov I, Puotinen M, Santos RS, Bates R (2012) Predictive modeling of dominant macroalgae abundance on temperate island shelves (Azores, Northeast Atlantic). In: Harris PT, Baker EK (eds) Seafloor geomorphology as benthic habitat: Geohab atlas of seafloor geomorphic features and benthic habitats. Elsevier Insights, London, p 169–182, doi: 10.1016/B978-0-12-385140-6.00008-6

Vandeperre F, Higgins R, Santos RS, Marcos C, Pérez-Ruzafa A (coords) (2008) Fishery regimes in Atlanto-Mediterranean European marine protected areas. EMPAFISH Project, Booklet no. WP2. Available at www.um.es/empafish/files/Deliverable%204-New%20Version.pdf

Vandeperre F, Higgins RM, Sánchez-Meca J, Maynou F and others (2011) Effects of no-take area size and age of marine protected areas on fishery yields: a meta-analytical approach. Fish Fish 12:412–426

Vanderklift MA, How J, Wernberg T, MacArthur LD, Heck KL Jr, Valentine JF (2007) Proximity to reef influences density of small predatory fishes, while type of seagrass influences intensity of their predation on crabs. Mar Ecol Prog Ser 340:235–243

Verlaque M (1990) Relationships between *Sarpa salpa* (L.) (Teleostei, Sparidae), other browser fishes, and the Mediterranean algal phytobenthos. Oceanol Acta 13: 373–388

Wakeling JM, Johnston IA (1998) Muscle power output limits fast-start performance in fish. J Exp Biol 201: 1505–1526

Wainwright PC, Bellwood DR, Westneat MW (2002) Eco-morphology of locomotion in labrid fishes. Environ Biol Fish 65:47–62

Wood SN (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc 99:673–686

Wood SN (2006) Generalized additive models: an introduction with R. Chapman & Hall/CRC, London

Wood SN, Augustin NH (2002) GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecol Model 157: 157–177

Zuur AF, Ieno EN, Smith GM (2007) Analysing ecological data. Springer, New York, NY

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York, NY

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 1:3–14

Zuur AF, Schmiing M, Tempera F, Afonso P, Saveliev AA (2012) Zero inflated GLMs with spatial correlation—analysis of parrotfish abundance. In: Zuur AF, Saveliev AA, Ieno EN (eds) Zero inflated models and generalized linear mixed models with R. Highland Statistics, Newburgh, p 231–256

Appendix

Table A1. Model validation and generality of final generalised additive models for 11 coastal reef fishes in the Azores showing the Spearman's rank correlation (ρ) of predicted and observed values from the model fitting and validation data set, rank sum statistics from the Wilcoxon signed rank test (V), percentage of correct predicted presences and absences (CP), and the area under the curve (AUC, for binomial models). ***p < 0.001, **p < 0.01, *p < 0.05

Species	ρ		V (Validation)	CP (%) (Validation)	AUC (Fitting)
	Fitting	Validation			
Coris julis[a]	0.81***	0.78***	2961	89	–
Labrus bergylta	0.57***	0.54**	2703	77	–
Serranus atricauda	0.67***	0.78***	3114	76	–
Sphyraena viridensis	0.56***	0.48***	1530***	88	0.92
Symphodus caeruleus	0.56***	0.40***	2327	75	0.86
Sarpa salpa	0.61***	0.62***	2334.5	74	–
Sparisoma cretense	0.65***	0.49***	2175*	77	–
Abudefduf luridus	0.76***	0.70***	2064**	65	–
Diplodus sargus	0.62***	0.46***	2494	95	–
Pagellus spp.	0.41***	0.51***	2783	78	0.81
Trachurus picturatus	0.44***	0.56***	3093	76	0.78

[a]Significant difference between model fitting and validation data set (Mann-Whitney test)