Habitats of ten demersal species in the Gulf of Lions and potential implications for spatial management

Marie Morfin[1,*], Nicolas Bez[2], Jean-Marc Fromentin[1]

[1]UMR 212 EME, IFREMER (Institut Français de Recherche pour l'Exploitation de la MER), 34200 Sète, France
[2]UMR 212 EME, IRD (Institut de Recherche pour le Développement), 34200 Sète, France

ABSTRACT: An improved knowledge of habitat utilization by demersal species is a pre-requisite for their spatial management. Based on scientific survey data collected over the period 1994–2010, the present study investigates relationships between 4 environmental factors and 10 demersal species in the Gulf of Lions (northwestern Mediterranean Sea). Generalized linear models provided statistically satisfying results in terms of both model explanatory and predictive powers. The 'biological zone' factors, based on the percentage of light penetration to the sea bottom and bottom temperature, were the most important factors, while sediments and benthic macrofauna were only significant for a few species. The type of associations varied among species, resulting in different spatial predictions among species. The spatial structures of species distributions appeared to be due more to habitat preferences that are spatially auto-correlated than to intra-specific population dynamics. The use of a spatial optimization procedure on the predicted species distributions allowed the detection of a set of 7 zones covering 17% of the studied region, that included at least 20% of each species' abundance and that reflected the diversity of the species' habitats. This preliminary result illustrates the potential of further analyses on Marine Protected Areas as a tool for the conservation of the demersal community in this region.

KEY WORDS: Fish habitat · Marine living resources · Mediterranean · Generalized linear models · Model selection · Marine Protected Area · MARXAN

INTRODUCTION

The role of Marine Protected Areas (MPAs) is to serve conservation purposes and maximize ecological services (Freire & Garcia-Allut 2000, Apostolaki et al. 2002). However, MPAs are often designed without adequate information on how the ecosystem works, thus failing to meet the objectives for which they were implemented (Roberts et al. 2003). Modelling spatial distributions of the exploited species can help stakeholders and managers to decide where to place MPAs and the type of management strategies that are most viable for the species of concern, e.g. protection of certain life stages of specific species or overall enhancement of multiple fisheries (Sumaila et al. 2007, Elith & Leathwick 2009). An accurate knowledge of fish ecological niches is also useful to predict the spatial distributions of target species and can help to better understand the effects of changing environmental conditions on these distributions (Guisan & Thuiller 2005, Ramírez-Bastida et al. 2008). Habitat models have been widely used for either terrestrial or aquatic species thanks to development of statistical tools, such as generalized linear models (GLMs), generalized additive models (GAMs) and random forests and boosted regression trees (Guisan & Zimmermann 2000, Breiman 2001, Elith & Leathwick 2009). In particular, GLMs have been used successfully to predict the mean response of various species to environmental factors, and are

flexible as they can be applied to data that are not necessarily normally distributed, such as those collected during scientific trawl surveys (McCullagh & Nelder 1989).

This study focuses on the Gulf of Lions, located along the French coast in the northwestern Mediterranean Sea, a marine biodiversity hotspot that is subject to various anthropogenic pressures, such as pollution, fishing and climate changes (Bethoux et al. 1990, Caddy et al. 1995, Bianchi & Morri 2000, Myers et al. 2000, Mittermeier et al. 2005, Coll et al. 2010). A wide continental shelf extends to 200 m depth and ends with an abrupt slope eroded by several submarine canyons. This area is one of the most productive in the Mediterranean Sea owing to a number of hydrographic features, including the Rhone River run-off, frontal activities due to the geostrophic circulation and wind-driven upwelling cells (Millot 1990). The continental slope and submarine canyons may constitute a refuge from exploitation for large individuals of several commercial species, as these areas are less accessible to bottom trawlers. This may explain why the Gulf of Lions is still the host of a high level of biodiversity, despite the intense fishing effort there since the mid-20th century, resulting in growth overexploitation of several species, such as hake *Merluccius merluccius*, red mullet *Mullus barbatus* and horned octopus *Eledone cirrhosa* (Aldebert 1997, Papaconstantinou & Farrugio 2000, Gaertner et al. 2007, GFCM 2011).

Previous analyses in the Gulf of Lions highlighted that demersal fish communities are distributed mainly along a depth gradient (from the coastal zone to the upper slope) and also along a longitudinal gradient, in association with benthic macrofauna and substratum (Gaertner et al. 1999, 2002, Gaertner 2000). These studies, however, employed multivariate techniques, which do not include species distribution modelling.

The present study investigated the habitat of 10 demersal species in the Gulf of Lions and predicted their spatial distribution in abundance using a GLM approach incorporating 4 environmental variables (temperature, sediment, light penetration and benthos macrofauna) as explanatory factors (McCullagh & Nelder 1989). The inter-annual stability of the spatial distributions of the 10 species in May and June (Morfin et al. 2012) makes it relevant to investigate their temporally persistent habitats. From predicted maps of species densities (ind. km^{-2}), we examined the possibility of defining a MPA for the 10 species, using a MARXAN procedure (Ball & Possingham 2000). MARXAN is one of the most widely used soft-ware products for conservation planning. It is a stochastic optimization procedure that identifies, from a set of candidate sites, a subset of sites which are characterized by some particular biodiversity features, while attempting to minimize the cost of their inclusion in an MPA (McDonnell et al. 2002). This cost can be related to social, economical, or ecological features, or a combination thereof.

MATERIALS AND METHODS

Species data

Species data were collected during the International Bottom Trawl Survey in the Mediterranean Sea (MEDITS) conducted every year in May and June from 1994 to 2010. For the Gulf of Lions, 65 fixed sampling stations that have been sampled since 1994 were chosen, using a random stratified design (see Fig. 1 and Bertrand et al. 2002). Trawling was carried out for periods of 30 min (to 200 m depth) or 1 h (from 200 m depth) in daylight following a standardized protocol (Fiorentini et al. 1999). Catchability and accessibility were assumed to be constant over the whole area, though the sites located on the slope were known to be less accessible to the trawl, while adults of several species, like *Merluccius merluccius*, are more abundant in deeper areas. To counterbalance the lower accessibility of the trawl, the MEDITS Group doubled the sampling duration in deeper waters (>200 m; Bertrand et al. 2002 for more details). Each catch content was sorted by species, counted and weighed.

Since the inception of the MEDITS survey, 300 different species have been identified in the Gulf of Lions. However, many are rare or low in abundance, which makes a habitat model irrelevant for them (Mérigot et al. 2007). Therefore, we only considered the species present in at least 20% of the hauls and that are properly selected by the fishing gear. Furthermore, we focused on exploited species and selected species in different trophic and taxonomic (decapods, cephalopods, elasmobranchs, fish) positions (Bănaru et al. 2012, Morfin et al. 2012). This selection led us to retain the following 10 species: European hake *Merluccius merluccius*, Atlantic horse mackerel *Trachurus trachurus*, Mediterranean horse mackerel *Trachurus mediterraneus*, grey gurnard *Eutrigla gurnardus*, red gurnard *Aspitrigla cuculus*, Norway lobster *Nephrops norvegicus*, red mullet *Mullus barbatus*, horned octopus *Eledone cirrhosa*, small-spotted catshark *Scyliorhinus canic-*
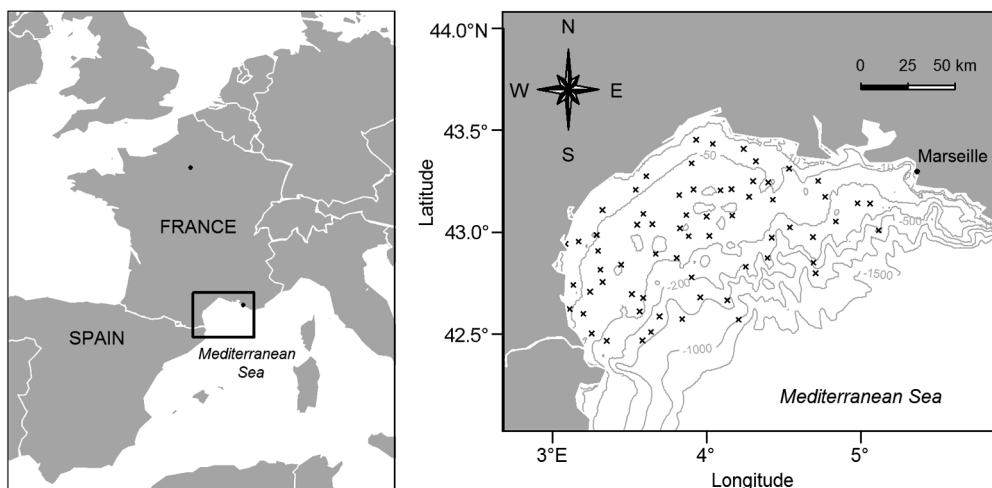
Fig. 1. Study area and sampling sites. Map of the Gulf of Lions and the 65 sampling sites (identified by crosses) during the whole MEDITS survey (1994–2010). The positions of the sites were set by a stratified sampling scheme, according to bathymetry (contour lines in grey, in meters)

ula and elegant cuttlefish *Sepia elegans*. Although both *Trachurus* species are pelagic, we considered that they were properly sampled by the MEDITS gear, as their distributions matched those observed during the pelagic acoustic survey PELMED, carried out in the same region.

Environmental variables

Following previous studies on fish habitat and data availability, we considered the following 5 environmental variables: depth, biological zones, bottom temperature, sediment type and benthos type (Johnson et al. 2013).

Depth data were obtained by a digital elevation model from data provided by the SHOM (Service Hydrographique et Océanographique de la Marine). In the study area the depth distribution is highly skewed, owing to the presence of the wide continental shelf and the abrupt slope (Fig. 2). As a first trial, a logarithm transformation combined with polynomial terms appeared then necessary to establish a relationship with species distributions. However, predictions out of the range of observed depths were very sensitive to the polynomial fit. They were also very uncertain in the shelf-edge area where very few large depth observations were made. While bathymetry is one of the most used factors in demersal fish niche models, the causal link between depth and demersal species is hardly a direct one. Depth is more likely the driving factor of many other biologi-

cal and physical processes which directly contribute to define demersal fish habitat.

We thus considered biological zones as an alternative to bathymetry. Biological zones were modelled by the EuseaMap project and defined by the percentage of light penetration at the bottom, as a function of bathymetry and turbidity (Cameron & Askew 2011). The 'infra-littoral' zone was allocated to values of up to 1% of light penetration, the 'circa-littoral' zone to values between 1 and 5%, and the 'bathyal' zone to values >5%. These thresholds (1 and 5%) are known to impact photosynthetic activity and the presence of different kinds of algae that induce different biotopes (Cameron & Askew 2011). From a statistical point of view, the relationships with species distributions will be easier to establish as the coverage of the different zones is balanced in the study area.

Bottom temperature was measured *in situ* during the survey by an onboard SCANMAR device from 1996 to 2010. We used a variance decomposition method (empirical orthogonal functions) to demonstrate the high inter-annual temporal stability of bottom temperature. The averaged values over the time period represented 78.2% of the total variance of the bottom temperature (for details see Fig. S1 in the Supplement at www.int-res.com/articles/suppl/m547p219_supp.pdf). These results legitimate the use, in this study, of averaged values of the bottom temperature over the whole period at each site.

A seabed sediment map was obtained from the EuSeamap project, which collated sediment data from various sources and built a map at 250 m resolu-
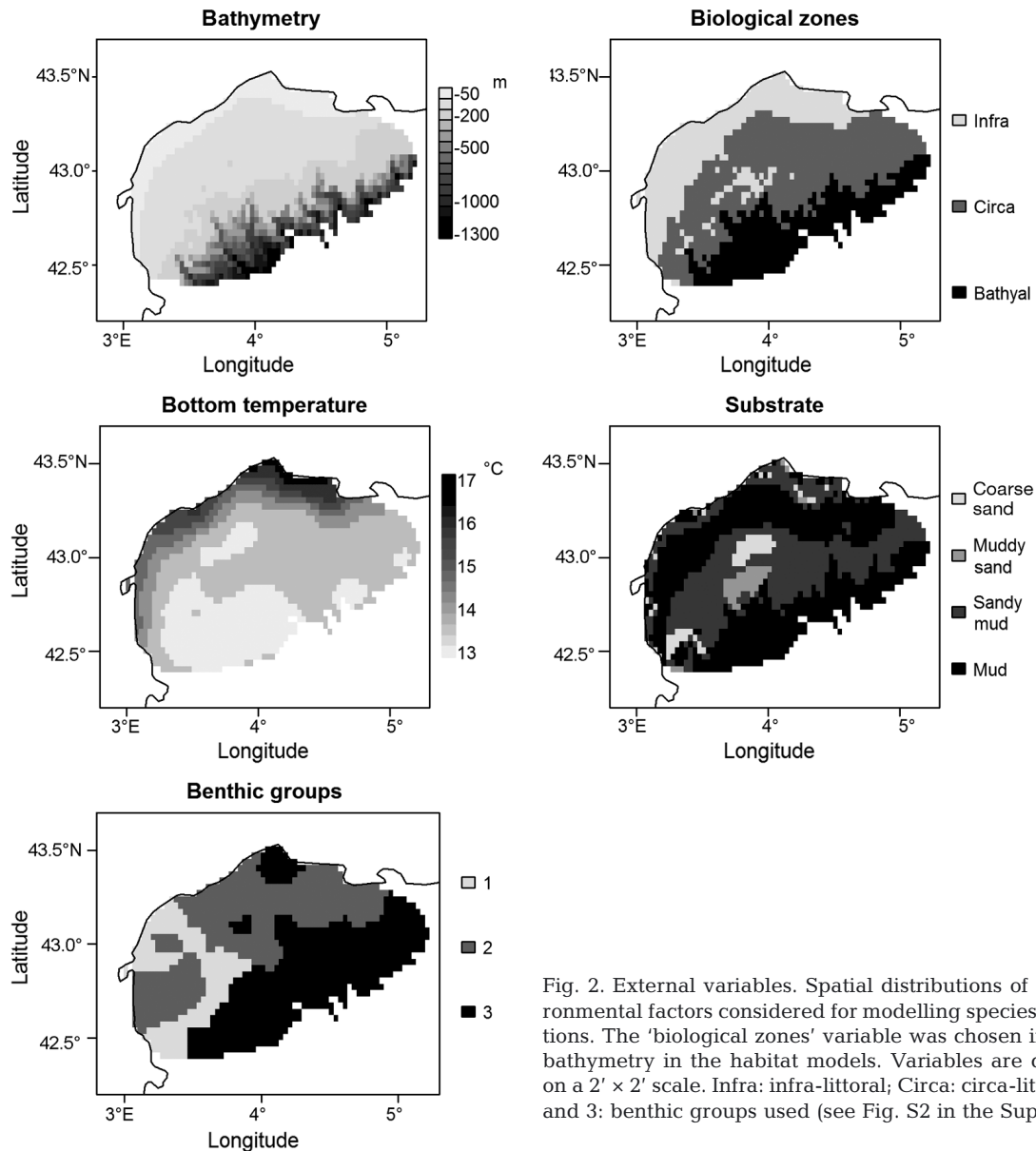
Fig. 2. External variables. Spatial distributions of the environmental factors considered for modelling species distributions. The 'biological zones' variable was chosen instead of bathymetry in the habitat models. Variables are displayed on a 2′ × 2′ scale. Infra: infra-littoral; Circa: circa-littoral; 1, 2 and 3: benthic groups used (see Fig. S2 in the Supplement)

tion, with 4 sediment types: 'gravel', 'sand', 'muddy-sand', 'sandy-mud' and 'mud' (Cameron & Askew 2011). As only one site is associated with the 'sand' class, gravel and sand classes were included in the same 'coarse sand' class for statistical purposes.

Benthos macrofauna is an important factor for the habitat of demersal fishes (see e.g Gaertner et al. 1999). Data on benthic species were collected by the MEDITS survey in 2000–2002, 2004 and 2006, and grouped into 15 categories. To reduce the number of classes, we performed a cluster analysis, which led to the retention of 3 main benthos groups. The 3 groups were dominated by mantis shrimp and tunicates, and mostly differed by their overall abundance, whatever the type of fauna, rather than by their species compo-

sition. Group 1 is the most abundant, while Group 3 is the least abundant. Nonetheless, Group 2 is the most abundant in sea urchins and polychaetes, and Group 3 is the most abundant in Cnidaria, crabs and crinoids (Fig. S2 in the Supplement).

Model formulation and calibration

Relationships between species count data and environmental variables were analysed following a GLM approach. In the present analysis, the retained variables of interest were the number of individuals per species and site summed over the 17 yr, as Morfin et al. (2012) have shown that the spatial distributions

of those species studied were highly stable over time. The corresponding trawled surface (in km^2) was accounted for in the model as an offset. For count data, Poisson and negative binomial (NB) distributions are natural choices. Still, for all species the Poisson distribution appeared to be clearly inappropriate according to the inspection of the residuals. We thus developed NB models with a logarithm link function to approach linearity and maintain model predictions within a range of values consistent with the original data (Guisan & Zimmermann 2000). Nonetheless, non-linearities were detected between the log expectancy of the response variable and the bottom temperature (the only continuous covariate). The bottom temperature (BT) was thus tested using first-, second- and third-order polynomial terms, to reflect a potential optimum (second degree) in the response of species abundance to temperature, and a skewed form of the response around the optimum (third degree). The interactions between temperature and biological zones were also tested, resulting in 92 candidate models.

To detect potential multicollinearities whose interaction and polynomial terms are often the source, the generalized variance inflation factor (GVIF) was used (Montgomery & Peck 1982, Fox & Monette 1992). As a rule of thumb, multicollinearity may be problematic (increasing the parameter variances and raising interpretation difficulties) when the GVIF$^{1/2df}$ is >2 (Fox & Monette 1992). This problem was solved for polynomial terms by orthogonalizing them, but multicollinearities were still found for interaction terms.

Models were implemented using R (R Development Core Team 2011). The 'glm.nb' function of the MASS library was used for adjustment and prediction (Venables & Ripley 1994).

Model selection

Two criteria were used to select models, a corrected Akaike information criterion (AICc) and a prediction performance index estimated by cross-validation. The necessary bias correction of the AIC for small sample size was estimated by a Gaussian approximation. In addition, AIC-based selection may result in overfitted models and/or in inappropriate promotion of complex hypotheses (Burnham & Anderson 2003). Cross-validation procedures have been proposed to overcome this problem and estimate how accurately a predictive model will perform in practice (Geisser 1975, Efron & Tibshi-

rani 1993). Though less robust than the 'k-fold' cross-validation procedure, leave-one-out (Loo) cross-validation was chosen to keep enough data in the 'training' sample (Arlot & Celisse 2010). Prediction performance was then measured by the percentage of mean absolute error (PMAE), i.e. MAE divided by mean abundance, on the data left out. A value >1 indicates that the average prediction error is higher than the average abundance. Both selection procedures, AICc and Loo, were applied for comparison purposes, but we used models selected by Loo for the species distribution predictions (see 'Model evaluation') as the criterion is based on prediction performance.

Model evaluation

Species distribution models were evaluated using residual analysis and deviance estimates. The standardized deviance residuals were used to check the model adequacy. Furthermore, plots of these residuals against fitted values and each explanatory variable allowed the identification of unexpected patterns in the deviance. The proportion of deviance explained by the predictors was also calculated to assess the explanatory power of the model.

To quantify and visualize the impact of the continuous explanatory bottom temperature variable, marginal effects were estimated by the average of abundance predictions from the selected model for several fixed values of this variable. The uncertainty around the fitted values of the response variable was estimated by a bootstrap procedure (Efron & Tibshirani 1993, McCullough 1994).

Species distribution maps

Predictions of species abundance were calculated from habitat models selected by the Loo procedure, as linear combinations of the explanatory variables on a 2′ × 2′ grid (Table 1). Species distributions over the Gulf of Lions were thus built using maps for each of the 4 predictor variables to predict each species' habitat (Fig. 2). The maps of benthos groups and bottom temperature were not available over the whole study area and were thus interpolated. Bottom temperature was interpolated on the prediction grid by ordinary kriging (Matheron 1963, Cressie 1993). As the benthos is a categorical variable, values were predicted using Voronoï polygons (each pixel was associated with the group of the nearest observed site).

Table 1. Summary of selected models for the 10 demersal species, using binomial negative regressions. For each species, first line shows results from the model selected by the leave-one-out (Loo) cross-validation procedure; the second line (in italics) shows results from the model selected according to AICc. Predictors are biological zone (BZ), bottom temperature (T), substrate (SUB) and benthos (BENT). Asterisks between predictors indicate interactions between model predictors. Ex Dev: percentages of deviance explained by the covariates; PMAE: percentage of mean absolute error calculated by cross-validation, an indicator of model prediction power

Species	Selected model (number of parameters)	Ex Dev (%)	PMAE (%)
Hake	$BZ + T + T^2 + BENT + BZ^*(T + T^2)$ (12)	59.3	33
(*Merluccius merluccius*)	*$BZ + T + SUB + BZ^*T$ (10)*	*61.5*	*46*
Atlantic horse mackerel	$BZ + T + BENT$ (7)	46.5	58
(*Trachurus trachurus*)	*$BZ + T + SUB + BZ^*T$ (10)*	*76.3*	*63*
Mediterranean horse mackerel	$BZ + T + SUB$ (8)	48.8	69
(*Trachurus mediterraneus*)	*$BZ + T + T^2 + T^3 + BZ^*(T + T^2 + T^3)$ (13)*	*72.2*	*72*
Grey gurnard	$BZ + T + T^2 + SUB + BENT + BZ^*(T + T^2)$ (15)	86.5	40
(*Eutrigla gurnardus*)	*$BZ + T + T^2 + BENT + BZ^*(T + T^2)$ (12)*	*85.2*	*40*
Red gurnard	$SUB + BENT$ (7)	31.6	115
(*Aspitrigla cuculus*)	*$BZ + T + BENT + BZ^*T$ (9)*	*54*	*1513*
Norway lobster	$BZ + T + T^2 + BENT + BZ^*(T + T^2)$ (12)	87.2	78
(*Nephrops norvegicus*)	*$BZ + T + T^3 + SUB + BZ^*T^3$ (12)*	*89.6*	*79*
Red mullet	$BZ + T + T^2 + SUB + BENT + BZ^*(T + T^2)$ (15)	81.1	64
(*Mullus barbatus*)	*$BZ + T^3 + SUB + BENT + BZ^*(T^3)$ (12)*	*79.2*	*72*
Horned octopus	$BZ + T + T^2 + SUB$ (9)	61.3	30
(*Eledone cirrhosa*)	*$BZ + T + T^2 + BZ^*(T + T^2)$ (9)*	*66.4*	*31*
Small-spotted catshark	$T + SUB + BENT$ (8)	46.7	75
(*Scyliorhinus canicula*)	*$T + T^2 + T^3 + SUB$ (8)*	*49.3*	*79*
Cuttlefish	$BZ + T + T^2 + T^3 + BENT$ (9)	76.5	47
(*Sepia elegans*)	*$BZ + T + T^2$ (6)*	*75.6*	*49*

MARXAN analysis

MARXAN is an optimization algorithm which implements an objective function to minimize, including a penalization term for not achieving the conservation target and the cost of the reserve. The planning units were created by dividing the study area into 2′ × 2′ squares, and the target objectives were species distribution maps predicted by GLMs. The biodiversity target was formulated to ensure that at least 20% of abundance of each species was represented in a protected area network. This threshold was proposed at the 2002 Earth Summit and advocated as the minimum amount of each habitat to be represented in marine reserves (IUCN World Parks Congress 2003). Reserve area was used as a surrogate for the cost of the planning units, based on the assumption that the larger the reserve, the more costly the implementation and management. As the optimal solution may be highly fragmented, a penalty for the total boundary length of the reserve (boundary length modifier) was included to get the best compromise between the total area of the conservation system and its compactness (Stewart & Possingham 2005).

MARXAN uses a stochastic optimization algorithm (simulated annealing), which enables the detection of approximated solutions within a reasonable amount of time. Depending on the spatial distribution of each species and the form of the cost function, near-optimal solutions can be more or less difficult to detect. To ensure that the algorithm finds stable solutions, we computed 1 million iterations and repeated the procedure for 500 different initial values. The number of times each planning unit was included in the resulting solutions among the 500 runs is a measure of how essential any particular unit is to forming a comprehensive system. Finally, we displayed the best solution and selection frequency among the 500 runs for the scenarios with and without a boundary length penalty.

RESULTS

Model evaluation and interpretation

For all species, graphics of deviance residuals versus fitted values did not display any special pattern and <5% of the values were found outside the 95%-

confidence interval (Fig. S3 in the Supplement at www.int-res.com/articles/suppl/m547p219_supp.pdf). Quantile-quantile plots displayed no significant departure of deviance residuals from normal distribution, except for red gurnard (Fig. S4 in the Supplement). Deviations at extremities were observed for several species; however, the normality of error is not a condition of GLM quality but simply a description of model behaviour. Variograms of deviance residuals presented some auto-correlation structure for 3 species: hake, catshark and octopus (Fig. S5 in the Supplement). For the 2 mackerels, the spatial structure was only present with the model selected by Loo.

Several differences were detected between models selected by Loo and AICc. However, no general rule can be established, except that for the same level of complexity the factors selected by AICc explained slightly higher percentages of deviance (on average 67.4 and 70.9% respectively; Table 1). Overall the percentages of explained deviance were substantial, ranging from 31.6% (for red gurnard) to 89.6% (for the Norway lobster) and correlated with the number of parameters, varying from 6 to 15. Although no significant linear correlation was detected between covariates, there were some redundancies that make it impossible to distinguish the part of deviance explained by each covariate. That is why the sum of deviance explained by each covariate separately may be much greater than the total explained deviance, e.g. Norway lobster is strongly associated with the 2 biological zones and the bottom temperature (Table 2). However, these results highlight that the factors that were not selected by both selection procedures generally explained very low percentages of deviance, except in 2 cases. Depending on the model selection, a strong association was established between red gurnard distribution and temperature or substrate type and between Norway lobster

and benthos type and substratum. For a given species, the factors selected by both procedures displayed the same marginal effects. We thus focused on these factors for model interpretation.

The biological zone variable was systematically selected, except for small-spotted catshark and red gurnard. This factor explained 31.2% of species distribution deviance on average, with a maximum of 80.8% for Norway lobster. Cuttlefish was strongly associated with the circa-littoral zone; Norway lobster with the bathyal zone. The 7 other species were associated with both circa-littoral and infra-litorral zones (Fig. S6 in the Supplement).

Apart from red gurnard, the bottom temperature was always selected in the models and also explained an important amount of the deviance: 36.3% on average and up to 69% for the Norway lobster. Densities of the 2 species of horse mackerel and cuttlefish increased linearly with temperature (Fig. 3). The densities of hake and small-spotted catshark displayed an optimum around 13.5°C, while horned octopus and grey gurnard displayed an optimum around 14°C (Fig. 3). For Norway lobster, densities were decreasing with temperature.

Sediment type explained 9.4% of the deviance on average, with a maximum of 43.1% for red mullet. It was selected for all species expecting cuttlefish. Red gurnard, red mullet and horned octopus were associated with coarse sand; catshark was associated with muddy sand bottom; hake and Atlantic horse mackerel were associated with sandy mud bottom; Norway lobster and Mediterranean horse mackerel were associated with muddy bottoms (Fig. S7 in the Supplement).

Benthos groups explained 6.7% of the deviance on average, with a maximum of 50.7% for Norway lobster. Group 2 was referential because species were ordered in abundance, though each group displayed some differences in species composition (see 'Materi-

Table 2. Proportion of deviance explained by each covariate included in both selected models (Loo/AICc). NA: covariate not included in the model. For scientific names of species see Table 1

| Species	Biological zones (BZ)	Temperature (T)	Substratum	Benthos	Interaction BZ*T
Hake	20.7/20.9	2.7/2.0	NA/9.4	1.8/NA	32.0/25.7
Atlantic horse mackerel	42.0/43.0	10.2/10.4	NA/3.7	3.4/NA	NA/27.4
Mediterranean horse mackerel	33.9/35.0	18.6/34.6	2.8/NA	NA	NA/25.7
Grey gurnard	19.3/23.7	3.0/3.7	2.0/NA	5.3/6.7	55.7/56.0
Red gurnard	NA/5.0	NA/14.2	30.0/NA	14.7/15.0	NA/16.1
Norway lobster	70.8/71.6	56.2/63.0	NA/39.8	50.7/NA	8.2/6.0
Red mullet	14.8/14.7	1.3/0.3	42.8/43.1	11.6/11.7	10.1/22.5
Horned octopus	37.9/38.3	22.5/22.7	4.7/NA	NA	NA/6.4
Small-spotted catshark	NA	35.1/41.0	8.8	10.0/NA	NA
Cuttlefish	65.8	48.6/44.8	NA	3.5/NA	NA
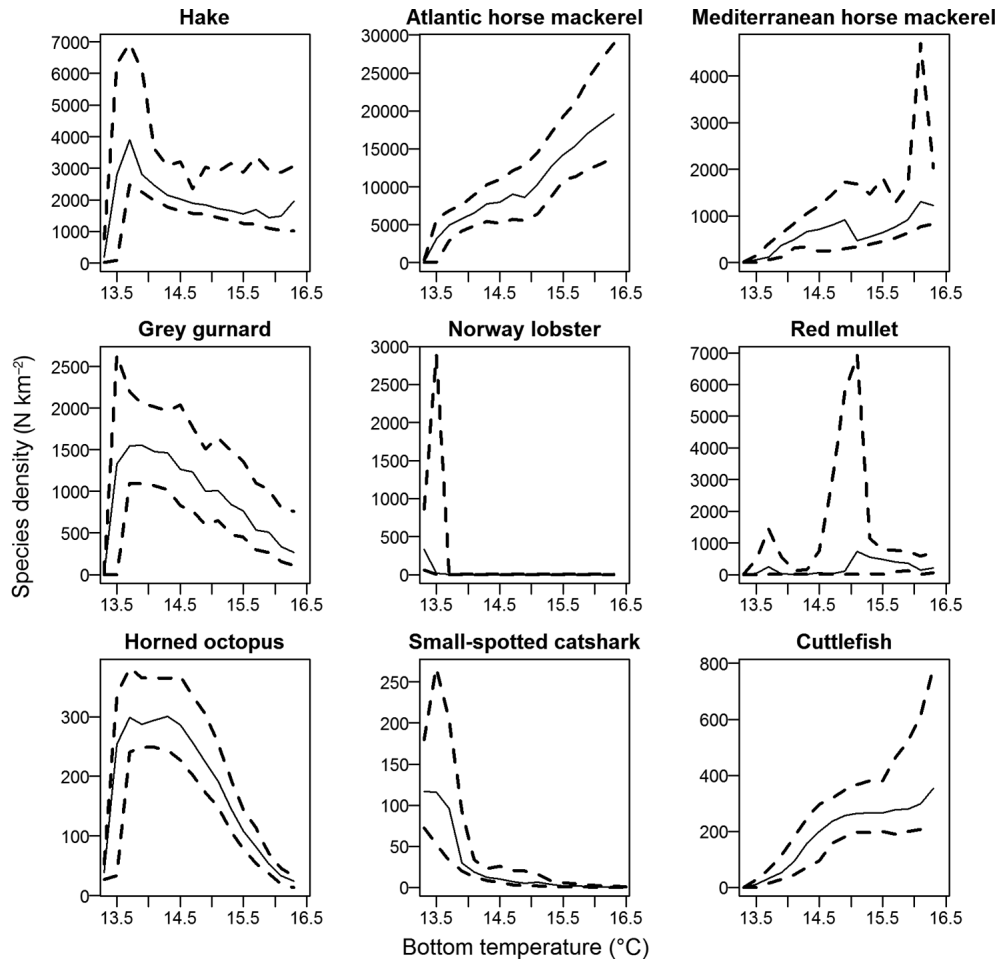
Fig. 3. Marginal response in species density depending on bottom temperature, as predicted by the model selected by the Loo procedure. Red gurnard case is not shown as the Loo selection procedure did not select bottom temperature for this species. Distributions of the marginal response were estimated by bootstrap; solid lines display the means and dashed lines the 10 and 90% quantiles. For scientific names of species see Table 1

als and methods'). Red mullet, red gurnard and catshark were associated with low benthos abundance level; grey gurnard and Atlantic horse mackerel, to an intermediate level; and hake and Norway lobster, to a high level (Fig. S8 in the Supplement).

The predictive power of models was measured by PMAE, a value >1 indicating a poor predictive capacity. Except for red gurnard (PMAE = 115%), the 9 other species models selected by Loo showed satisfactory predictive performance (PMAE = 30–75%; Table 1). The species distribution maps were thus predicted using the models selected by Loo.

Species distribution maps

Mapping model predictions over the Gulf of Lions highlighted, as expected, different spatial patterns across species (Fig. S9 in the Supplement). Species found close to the coasts were horse mackerels, cuttlefish and red mullet. Atlantic horse mackerel was also found on the west side of the slope; and red mul-

let in the west central part of the shelf. In contrast, Norway lobster was distributed on the eastern side of the shelf and over the whole slope (the greatest predicted values being obtained in the western sector of the slope). Hake, grey gurnard and octopus were mostly found on the central shelf, octopus being more abundant on the western side, while hake was more abundant on the eastern side.

Prediction uncertainty was measured by the coefficient of variation (CV) of predictions, a value >1 indicating considerable uncertainty in prediction. CV values were generally <1, with median values ranging from 0.16 (for horned octopus) to 0.71 (for Norway lobster), reflecting overall good predictions. Furthermore, CV values were consistent with the index of prediction performance, PMAE, as median values of CV decreased with increasing PMAE (Table 1). The poorest predictions (CV > 1) appeared to be due to the values of explanatory variables outside the range used to calibrate the models. These values covered 2% of the 10 maps and were mostly detected on the eastern side of the slope and on the western side

of the shelf, where temperature was not directly measured, but interpolated by kriging. In those areas, kriged values were significantly higher than observed values. The predictions of which CV values were >1 were discarded for MARXAN analysis.

MARXAN analysis

Among 500 runs using different initializing values, we looked at the best solution and the frequency of selection of each planning unit (note that the penalty for not achieving the conservation targets in the objective function was set to the minimum value where all conservation targets were met). Without any boundary length constraint, the set of planning units selected was highly scattered over the whole

area, reflecting both the sparsity of some species distributions, e.g. grey gurnard, and the differences between species distributions, e.g. Norway lobster versus cuttlefish, the former being offshore, while the latter is coastal (Fig. S9).

The weight allocated to the boundary length penalty (named boundary length modifier [BLM]) was set in order to get the best trade-off between the boundary length and the total area of the reserve system (Fig. S10 in the Supplement). The more optimal planning design solution according to the BLM (0.25) occupied 16.4% of the study region and was composed of 7 disjointed zones (Fig. 4, second line of the right column). This solution includes the 3.2% of planning units selected >80% of the time and the 67% selected >50% of the time. Among the 7 zones, 1 located on the west coast was preferential habitat
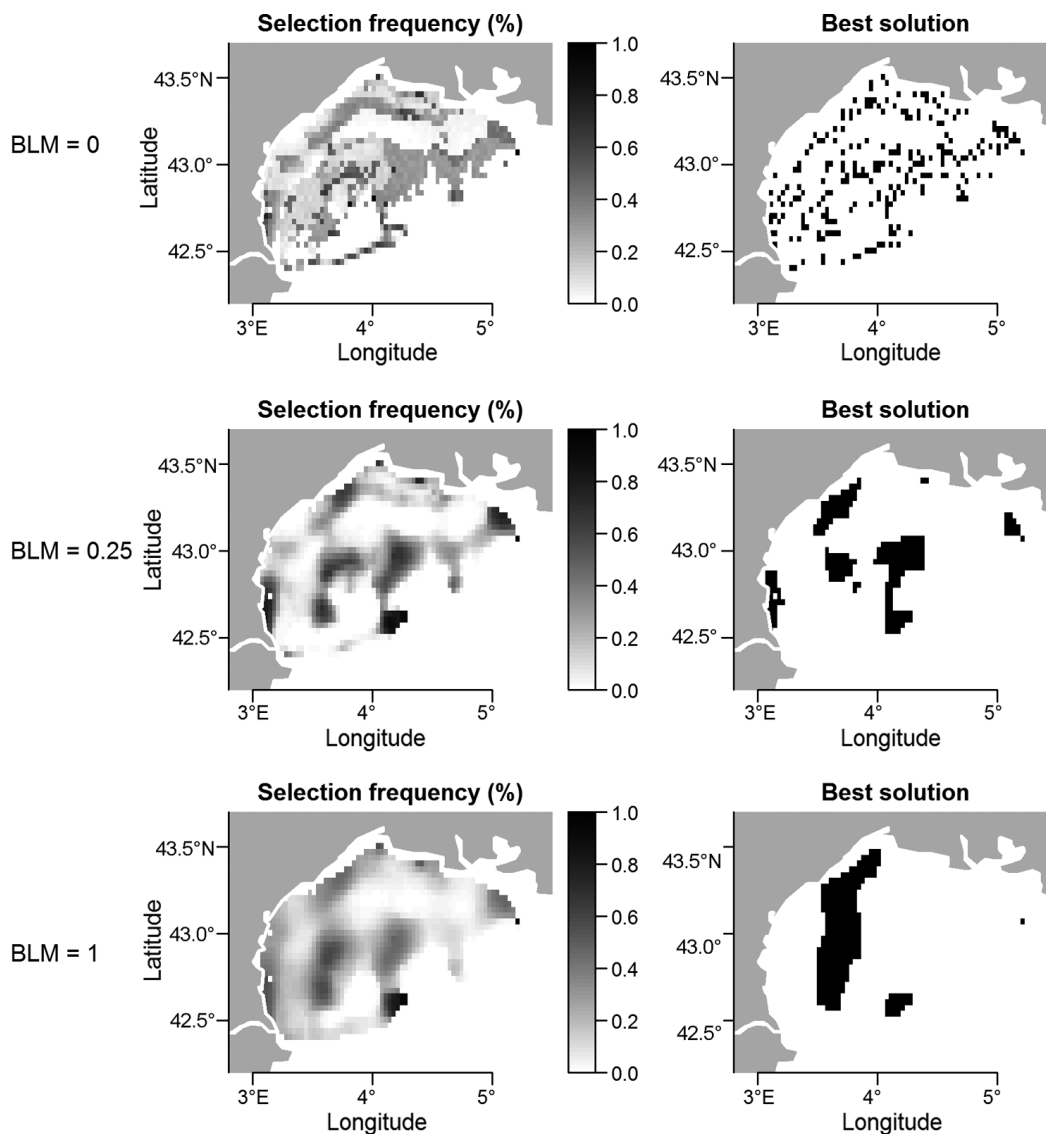

Fig. 4. Marine Protected Area design MARXAN outputs, i.e. the minimum set of areas containing 20% of each species abundance. Results according to 3 different compacity constraints are displayed: BLM (boundary length modifier) = 0 corresponds to no compacity constraint at all; BLM = 0.25 is the best trade-off between area and boundary length; BLM = 1 is the most compact solution whose area covered less than 20% of the study area (Fig. S4 in the Supplement). For each BLM value, the best solution (right panels) and the selection frequency of each planning unit (left panels) among 500 runs are displayed

for the 2 mackerels, grey gurnard, horned octopus and cuttlefish, while, on the middle of the coast, only both mackerels and cuttlefish were found. The tiny coastal zone a little further east was dominated by cuttlefish, red mullet and Mediterranean horse mackerel. The last zone on the coast is the eastern area, where hake and horned octopus were abundant. The 2 zones in the middle of the shelf corresponding to sandy bottom patches were particularly important, as all species except the 2 mackerels and cuttlefish were found in abundance (Fig. 2). The last zone located on the slope was the only area where the most preferential habitat of Norway lobster and some other species overlapped. Furthermore, they included the main types of physical habitats present in the Gulf of Lions in similar proportions, i.e. all the biological zones, the full range of bottom temperatures, all the sediment types and the 3 benthos groups.

DISCUSSION

In this study we characterized habitats and predicted distributions for 10 demersal species in the Gulf of Lions, using a GLM approach. The selected species distribution models showed good explanatory power, as the percentages of explained deviance were high for all species. Their predictive capabilities were also satisfying, except for red gurnard. For this species, the model performance was probably affected by the fact that its highest abundance was observed at the only sandy site. This strong association could not be clearly detected, as we had to combine this site with gravel bottom sites in the same 'coarse-sand' category, for statistical significance.

We compared the results from 2 different model selection procedures, Loo cross-validation and AIC. Overall, results were quite similar. In theory, the former should be preferred for predictive models, while the latter should be preferred for explanatory models. Still, there is no rule of thumb to choose a model selection procedure, and our goal was to highlight the uncertainty that it may involve. Unexpectedly, the AIC-based procedure did not systematically select models with a higher explanatory power than the cross-validation procedure. While we tested several degrees of polynomials for temperature, we kept fixed the number of classes for categorical variables. Thus, models including 4 classes of sediment type were systematically more penalized than the others. Thus any interpretation of identified habitat factors and the exploitation of predicted maps should

be treated with caution if these aspects were investigated.

While stable spatial auto-correlation patterns were previously observed for these 10 species, for 7 of them (Atlantic and Mediterranean mackerel, red and grey gurnard, red mullet, Norway lobster and cuttlefish), no spatial structure was found in the residual of the fitted models (Morfin et al. 2012). Such a result suggests that auto-correlation patterns found for those species were probably due to habitat preferences that are spatially auto-correlated rather than intra-specific population dynamics. The 3 other species (hake, catshark and horned octopus) were those for which models displayed the lowest percentages of explained deviance, suggesting that the remaining auto-correlation in the residuals may be explained by external factors not included in the model. As this is not straightforward to implement for binomial negative distribution, the residual spatial auto-correlation has not been handled by the model. For these species, model outputs should be interpreted with greater care, as they may bias the influence of other factors.

Searching for significant relationships between species abundance and measured habitat variables has been criticized for ascribing coincidental correlations or indirect relationships to direct causal links (Guisan & Thuiller 2005). However, insufficient knowledge still remains on factors influencing marine species, which makes it difficult to test and validate prior assumptions about causal relationships. Consequently, correlative approaches that make few or even no prior assumptions about underlying causal relationships are considered legitimate when attempting to understand the complex interactions between fish populations and their environment (Valavanis et al. 2004). Those correlations can in turn be used as the basis for subsequent hypothesis-driven studies aiming to determine demersal fish habitat requirements. However, model outputs must be interpreted with caution when the sampling design cannot be totally controlled and balanced. For instance, the range 13–13.5°C was only observed in the bathyal zone, where the relationship between temperature and the log of species abundance was positive for several species, while it was negative in other ranges. In this case, interactions may be only statistics and reflect the non-linearity of the relationship between temperature and the log of the response variable, instead of a variation in the effect of temperature depending on biological zones.

Geographical predictions using GLM require that explanatory variables are known over the whole

area. In practice, this is rarely the case, and explanatory variables need to be previously predicted over the whole area, as we did for temperature by kriging. This is another source of uncertainty which is not integrated in species predictions. For example, bottom temperature was slightly overestimated in the bathyal zone, resulting in spurious high densities for Atlantic horse mackerel and small-spotted catshark. However, the distribution of the same 10 fish species performed by kriging (i.e. by direct spatial interpolation of the abundance, see Morfin et al. 2012) are very similar as those obtained in the present study (i.e. through habitat modeling). Such a result is primordial, as it validates the pertinence of our approach and the external factors chosen.

Biological zone and bottom temperature were the main factors explaining species distributions. These factors are strongly related to depth, which has often been reported to be the main gradient along which faunal changes occur when studying shelf and upper-slope demersal assemblages (Johnson et al. 2012). In the Bay of Biscay and the Celtic Sea, juveniles of red gurnard and hake were primarily associated with bathymetry and secondarily with bottom temperature and salinity (Persohn et al. 2009). Juveniles of many demersal species occur predominantly within the inshore soft bottoms along the coast (Bartolino et al. 2008, Carlucci et al. 2009), where some ecological processes that enhance their survival take place (Kaiser et al. 1999). Accordingly, much of the essential marine fish habitat is in shallow coastal waters, even though some deep habitats such as rocky submarine canyons may constitute natural refuges for large individuals of demersal species (Yoklavich et al. 2000). However, the ecological relevance of the bathymetry is not demonstrated for these species, as associations with depth may hide preferences for other physical factors or prey availability (Murawski & Finn 1988). In our case, bottom temperature is not necessarily a proxy for depth, as it was not correlated with bathymetry inside each biological zone. We used biological zones rather than bathymetry, as the distribution of the latter is highly skewed in the Gulf of Lions, which made difficult to model it and led to less satisfactory models (according to both model selection criteria) than the models using biological zones.

These results also revealed substrate and benthos as substantial drivers of demersal species distributions. The benthos variable was introduced because it constitutes the base diet of most of the species considered, but it may also be a proxy for other processes, as benthic macrofauna are more sensitive to some environmental factors (e.g. depth, sediment type, salinity, pollution,…) than demersal fishes (Ferraro & Cole 2007, Nicolas et al. 2007). As some of these variables have already been introduced into the model, the interpretation of this factor is not straightforward.

The high percentages of explained deviance in most habitat models indicate that the 4 habitat factors used in this study were sufficient to explain most species distributions. However, some additional factors could be included in our models, such as salinity, organic carbon flux, prey resources, pollution and fishing pressure (Sanchez-Vidal et al. 2009, Coll et al. 2012, Johnson et al. 2012). As the species considered here are all generalist feeders and it is not possible to include overall potential prey abundance as a covariate in such models, prey availability should be measured by some indicators (Quéro & Vayne 1997). Data on other factors were not yet available at fine resolution during our study.

Distributions of the species studied here were highly stable over the whole period from 1994 to 2010 (Morfin et al. 2012). This makes relevant the present study goal of determining species habitat that persisted over time. The strong association demonstrated in this analysis between averaged species abundance and temporal rather stable external factors (substrate, bottom temperature and biological zones) would certainly still be important in habitat models, including temporal variability. Despite strong temporal stability, Morfin et al. (2012) also documented a positive relationship between species occupancy area and total abundance over 17 yr. According to the MacCall basin theory, habitat selection is density dependent to some degree. At low abundance, individuals occupy the most suitable habitat, whereas they are expected to spread to marginal (sub-optimal) habitats when abundance increases because of intraspecific competition for food and/or space (Fretwell & Lucas 1969, MacCall 1990). The results of the present paper demonstrate that the spatial distributions of species are strongly associated with environmental factors. Furthermore, these associations were established from data averaged over the time period and are thus independent of temporally variable habitats. Therefore, it is more likely that the spatial expansion of species at high abundance is due to a density-dependent process, as expected under the MacCall basin theory, rather than to spatio-temporal variations in some key environmental variables. Therefore, we can expect that the deviance explained by these factors in temporal models would be lower if the density-dependent process were not taken into account.

We performed a first quantitative analysis to investigate the potential relevance of an MPA in the Gulf of Lions, using MARXAN procedure (Ball & Possingham 2000). Our goal was to determine the minimum set of areas containing at least 20% of the abundance of each of the 10 target species, which is obviously more complicated than for one given species (Fromentin & Lopuszanski 2014). The first solution obtained without any boundary length constraint was a set of areas highly scattered over the whole Gulf of Lions, reflecting the sparsity of some species distributions as well as the differences between species distributions. Such a solution was obviously not operational in terms of management. Once increasing the compactness of the protected areas by increasing boundary length, we obtained more 'manageable' solutions. The design realising the best compromise between total area and compactness covered 15.6% of the study region and was comprised of 7 disjointed zones. As this design was still very sparse, we also considered a higher compactness constraint, which produced 3 distinct zones covering 17% of the study area (Fig. 4, third line of the right column). The importance of coverage is partly due to the horned octopus, whose minimum area (including 20% of its distribution) alone represents 12% of the study area. The high coverage appears to be due more to the sparsity of this species than to the variability between species distributions. The conservation objective should probably be lowered, as this region is highly exploited and closing 20% of the fishing grounds would probably be regarded as unacceptable by some stakeholders. Some choices and compromises at the ecological/scientific level, but also at economic and political levels, would be required (Sumaila et al. 2007, Yates & Schoeman 2013). Although this is beyond the scope of the present study, assessing the potential benefits of a MPA should also consider the effects of the redistribution of fishing effort outside the MPA and the spatial dynamics of the main species (Apostolaki et al. 2002).

As the surveys were only carried out in May and June, this study can hardly tackle seasonal issues. This period was selected as it is a recruitment period for many species. This might be problematic, as all species, except Norway lobster, make ontogenic migrations, which may induce seasonal variability. For many species, juveniles migrate between coastal and bathyal zones according to seasons, as nurseries are located in both areas (GFCM 2010). Adults of hake, the most documented species, migrate to the middle of the shelf to spawn over the entire year, the spawning peak being in winter and early spring.

Nevertheless, too little information is available on species distributions in the Gulf of Lions during the other periods of the year. However, it would be interesting to perform the same analysis on juveniles and adults separately to identify nursery and spawning areas. The potential habitat data of young-of-the year individuals of hake determined by Druon et al. (2015) are consistent with our results (temperature range and bottom type). Colloca et al. (2015) investigated the overlap of several demersal species nurseries, using spatial models. Although the overlap covered an overly large area according to fishing activity, such an approach deserves further attention. The study of adult distributions requires models fitted to zero-inflated data, as adult specimens are observed at a very low proportion of sampled sites (Heilbron 1994).

Finally, this analysis indicates that designing an MPA network for several species of interest in the Gulf of Lions is not straightforward and deserves more dedicated investigations. Furthermore, several sources of uncertainties were highlighted through the analysis (model selection, predictions, optimization algorithm), which could lead to spurious conclusions. The Bayesian framework is well adapted to uncertainty propagation, and new advances in integrating map uncertainty in MPA design software are on their way (Carvalho et al. 2011, Kujala et al. 2013).

LITERATURE CITED

Aldebert Y (1997) Demersal resources of the Gulf of Lions (NW Mediterranean). Impact of exploitation on fish diversity. Vie Milieu 47:275–284

Apostolaki P, Milner-Gulland EJ, McAllister MK, Kirkwood GP (2002) Modelling the effects of establishing a marine reserve for mobile fish species. Can J Fish Aquat Sci 59: 405–415

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Statist Surv 4:40–79

Ball IR, Possingham HP (2000) Marine reserve design using spatially explicit annealing, a manual. University of Queensland, Brisbane

Bănaru D, Mellon-Duval C, Roos D, Bigot JL and others (2012) Trophic structure in the Gulf of Lions marine ecosystem (north-western Mediterranean Sea) and fishing impacts. J Mar Syst 111:45−68

Bartolino V, Ottavi A, Colloca F, Ardizzone GD, Stefánsson G (2008) Bathymetric preferences of juvenile European hake (Merluccius merluccius). ICES J Mar Sci 65: 963−969

Bertrand JA, Gil De Sola L, Papaconstantinou C, Relini G, Souplet A (2002) The general specifications of the MEDITS surveys. Sci Mar 66:9−17

Bethoux JP, Gentili B, Raunet J, Tailliez D (1990) Warming trend in the western Mediterranean deep water. Nature 347:660−662

Bianchi CN, Morri C (2000) Marine biodiversity of the Mediterranean Sea: situation, problems and prospects for future research. Mar Pollut Bull 40:367−376

Breiman L (2001) Random forests. Mach Learn 45:5−32

Burnham KP, Anderson DR (2003) Model selection and multi-model inference: a practical information-theoretic approach, 2nd edn (2002), corrected 3rd printing. Springer, Heidelberg

Caddy JF, Refk R, Do-Chi T (1995) Productivity estimates for the Mediterranean: evidence of accelerating ecological change. Ocean Coast Manage 26:1−18

Cameron A, Askew N (2011) EUSeaMap — Preparatory action for development and assessment of a European broad-scale seabed habitat map, final report. Available at http://jncc.gov.uk/euseamap

Carlucci R, Giuseppe L, Porzia M, Francesca C and others (2009) Nursery areas of red mullet (Mullus barbatus), hake (Merluccius merluccius) and deep-water rose shrimp (Parapenaeus longirostris) in the eastern-central Mediterranean Sea. Estuar Coast Shelf Sci 83:529−538

Carvalho SB, Brito JC, Crespo EG, Watts ME, Possingham HP (2011) Conservation planning under climate change: toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time. Biol Conserv 144:2020−2030

Coll M, Piroddi C, Steenbeek J, Kaschner K and others (2010) The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PLoS ONE 5:e11842

Coll M, Piroddi C, Albouy C, Ben Rais Lasram F and others (2012) The Mediterranean Sea under siege: spatial overlap between marine biodiversity, cumulative threats and marine reserves. Glob Ecol Biogeogr 21:465−480

Colloca F, Garofalo G, Bitetto I, Facchini MT and others (2015) The seascape of demersal fish nursery areas in the North Mediterranean Sea, a first step towards the implementation of spatial planning for trawl fisheries. PLoS ONE 10:e0119590

Cressie N (1993) Statistics for spatial data, revised edn. Wiley-Interscience, New York, NY

Druon JN, Fiorentino F, Murenu M, Knittweis L and others (2015) Modelling of European hake nurseries in the Mediterranean Sea: an ecological niche approach. Prog Oceanogr 130:188−204

Efron B, Tibshirani R (1993) An introduction to the bootstrap. CRC Press, Boca Raton, FL

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677−697

Ferraro SP, Cole FA (2007) Benthic macrofauna−habitat associations in Willapa Bay, Washington, USA. Estuar Coast Shelf Sci 71:491−507

Fiorentini L, Dremière LY, Leonori I, Sala A, Palumbo V (1999) Efficiency of the bottom trawl used for the Mediterranean international trawl survey (MEDITS). Aquat Living Resour 12:187−205

Fox J, Monette G (1992) Generalized collinearity diagnostics. J Am Stat Assoc 87:178−183

Freire J, Garcia-Allut A (2000) Socioeconomic and biological causes of management failures in European artisanal fisheries: the case of Galicia (NW Spain). Mar Policy 24: 375−384

Fretwell S, Lucas H (1969) On territorial behaviour and other factors influencing habitat distribution in birds. Acta Biotheor 19:16−36

Fromentin JM, Lopuszanski D (2014) Migration, residency, and homing of bluefin tuna in the western Mediterranean Sea. ICES J Mar Sci 71:510−518

Gaertner JC (2000) Seasonal organization patterns of demersal assemblages in the Gulf of Lions (north-western Mediterranean Sea). J Mar Biol Assoc UK 80: 777−783

Gaertner JC, Mazouni N, Sabatier R, Millet B (1999) Spatial structure and habitat associations of demersal assemblages in the Gulf of Lions: a multicompartmental approach. Mar Biol 135:199−208

Gaertner JC, Bertrand JA, Souplet A (2002) STATIS-CoA: a methodological solution to assess the spatio-temporal organization of species assemblages. Application to the demersal assemblages of the French Mediterranean Sea. Sci Mar 66:221−232

Gaertner JC, Bertrand JA, Relini G, Papaconstantinou C and others (2007) Spatial pattern in species richness of demersal fish assemblages on the continental shelf of the northern Mediterranean Sea: a multiscale analysis. Mar Ecol Prog Ser 341:191−203

Geisser S (1975) The predictive sample reuse method with applications. J Am Stat Assoc 70:320−328

GFCM (General Fisheries Commission for the Mediterranean) (2010) Report of the twelfth session of the Scientific Advisory Committee. FAO Fisheries and Aquaculture, Budva

GFCM (2011) Report of the thirteenth session of the Scientific Advisory Committee. FAO Fisheries and Aquaculture, Marseille

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 Modell 135:147−186

Heilbron DC (1994) Zero-altered and other regression models for count data with added zeros. Biom J 36:531−547

IUCN (International Union for the Conservation of Nature) World Parks Congress (2003) Vth IUCN World Parks Congress. IUCN, Durban

Johnson AF, Jenkins SR, Hiddink JG, Hinz H (2013) Linking temperate demersal fish species to habitat: scales, patterns and future directions. Fish Fish 14:256−280

Kaiser MJ, Rogers SI, Ellis JR (1999) Importance of benthic habitat complexity for demersal fish assemblages. In: Benaka LR (ed) American Fisheries Society Symposium 22. American Fisheries Society, Bethesda, MD, p 212−223

Kujala H, Moilanen A, Araújo MB, Cabeza M (2013) Conservation planning with uncertain climate change projections. PLoS ONE 8:e53315

MacCall AD (1990) Dynamic geography of marine fish populations. University of Washington Press, Seattle, WA

➤ Matheron G (1963) Principles of geostatistics. Econ Geol 58: 1246–1266

McCullagh P, Nelder JA (1989) Generalized linear models, 2ⁿᵈ edn. CRC Press, Boca Raton, FL

➤ McCullough BD (1994) Bootstrapping forecast intervals: an application to AR(p) models. J Forecast 13:51–66

➤ McDonnell MD, Possingham HP, Ball IR, Cousins EA (2002) Mathematical methods for spatially cohesive reserve design. Environ Model Assess 7:107–114

➤ Mérigot B, Bertrand JA, Mazouni N, Manté C, Durbec JP, Gaertner JC (2007) A multi-component analysis of species diversity of groundfish assemblages on the continental shelf of the Gulf of Lions (north-western Mediterranean Sea). Estuar Coast Shelf Sci 73:123–136

➤ Millot C (1990) The Gulf of Lions' hydrodynamics. Cont Shelf Res 10:885–894

Mittermeier RA, Gil PR, Hoffman M, Pilgrim J and others (2005) Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions. Conservation International, Washington, DC

Montgomery DC, Peck EA (1982) Introduction to linear regression analysis. Wiley, New York, NY

➤ Morfin M, Fromentin JM, Jadaud A, Bez N (2012) Spatio-temporal patterns of key exploited marine species in the northwestern Mediterranean Sea. PLoS ONE 7:e37907

➤ Murawski SA, Finn JT (1988) Biological bases for mixed-species fisheries: species co-distribution in relation to environmental and biotic variables. Can J Fish Aquat Sci 45:1720–1735

➤ Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858

➤ Nicolas D, Le Loc'h F, Desaunay Y, Hamon D, Blanchet A, Le Pape O (2007) Relationships between benthic macro-fauna and habitat suitability for juvenile common sole (*Solea solea*, L.) in the Vilaine estuary (Bay of Biscay, France) nursery ground. Estuar Coast Shelf Sci 73:639–650

➤ Papaconstantinou C, Farrugio H (2000) Fisheries in the Mediterranean. Mediterr Mar Sci 1:5–18

➤ Persohn C, Lorance P, Trenkel VM (2009) Habitat preferences of selected demersal fish species in the Bay of Biscay and Celtic Sea, North-East Atlantic. Fish Oceanogr 18:268–285

Quéro JC, Vayne JJ (1997) Les poissons de mer des pêches françaises. Delachaux & Niestle, Paris

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

➤ Ramírez-Bastida P, Navarro-Sigüenza AG, Peterson AT (2008) Aquatic bird distributions in Mexico: designing conservation approaches quantitatively. Biodivers Conserv 17:2525–2558

➤ Roberts CM, Andelman S, Branch G, Bustamante RH and others (2003) Ecological criteria for evaluating candidate sites for marine reserves. Ecol Appl 13:199–214

➤ Sanchez-Vidal A, Pasqual C, Kerhervé P, Heussner S and others (2009) Across margin export of organic matter by cascading events traced by stable isotopes, north-western Mediterranean Sea. Limnol Oceanogr 54:1488–1500

➤ Stewart RR, Possingham HP (2005) Efficiency, costs and trade-offs in marine reserve system design. Environ Model Assess 10:203–213

➤ Sumaila U, Zeller D, Watson R, Alder J, Pauly D (2007) Potential costs and benefits of marine reserves in the high seas. Mar Ecol Prog Ser 345:305–310

➤ Valavanis VD, Georgakarakos S, Kapantagakis A, Palialexis A, Katara I (2004) A GIS environmental modelling approach to essential fish habitat designation. Ecol Modell 178:417–427

Venables WN, Ripley BD (1994) Modern applied statistics with S-PLUS. Springer-Verlag, New York, NY

➤ Yates KL, Schoeman DS (2013) Spatial Access Priority Mapping (SAPM) with Fishers: a quantitative GIS method for participatory planning. PLoS ONE 8:e68424

Yoklavich M, Greene HG, Cailliet GM, Sullivan DE, Lea RN, Love MS (2000) Habitat associations of deep-water rockfishes in a submarine canyon: an example of a natural refuge. Fish Bull 98:625–641