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

Morfin Marie ^{1,*}, Bez Nicolas ², Fromentin Jean-Marc ¹

¹ IFREMER Inst Francais Rech Exploitat MER, UMR 212, EME, F-34200 Sete, France.

² IRD, UMR 212, EME, F-34200 Sete, France.

* Corresponding author : Marie Morfin, email address : marie.morfin@ifremer.fr

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.

Keywords : 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 in the first place (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

59 certain life stages of specific species or overall enhancement of multiple fisheries
60 (Sumaila et al. 2007, Elith & Leathwick 2009). An accurate understanding of fish
61 ecological niches is also useful to predict the spatial distributions of target species and
62 can help to better understand the effects of changing environmental conditions on these
63 distributions (Guisan & Thuiller 2005, Ramírez-Bastida et al. 2008). Habitats models
64 have been widely used for either terrestrial or aquatic species thanks to development of
65 statistical tools, such as Generalized Linear Models (GLMs), Generalized Additive
66 Models (GAMs), Random Forests and Boosted Regression Trees (Guisan &
67 Zimmermann 2000, Breiman 2001, Elith & Leathwick 2009). In particular, GLMs have
68 been used successfully to predict the mean response of various species to environmental
69 factors, and are flexible as they can be applied to data that are not necessarily normally
70 distributed, such as those collected during scientific trawl surveys (McCullagh & Nelder
71 1989).

72

73 This study focuses on the Gulf of Lions, located along the French coast in the
74 northwestern Mediterranean Sea, a marine biodiversity hotspot that is subject to various
75 anthropogenic pressures, such as pollution, fishing and climate changes (Bethoux et al.
76 1990, Caddy et al. 1995, Myers et al. 2000, Bianchi & Morri 2000, Mittermeier et al.
77 2005, Coll et al. 2010). A wide continental shelf extends until 200 m depth and ends by
78 an abrupt slope eroded by several submarine canyons. This area is one of the most
79 productive of the Mediterranean Sea owing to a number of hydrographic features,
80 including the Rhone river run-off, frontal activities due to the geostrophic circulation
81 and wind-driven upwellings cells (Milot 1990). The continental slope and submarine
82 canyons may constitute a refuge from exploitation for large individuals of several
83 commercial species, as these areas are less accessible to bottom trawlers. This may
84 explain that the Gulf of Lions is still the host for a high level of biodiversity, despite the
85 intense fishing effort since the mid-20th century, resulting in a growth overexploitation
86 for several species, such as hake (*Merluccius merluccius*), red mullet (*Mullus barbatus*)
87 and horned octopus (*Eledone cirrhosa*) (Aldebert 1997, Papaconstantinou & Farrugio
88 2000, Gaertner et al. 2007, GFCM 2011).

89

90 Previous analyses in the Gulf of Lions highlighted that demersal fish communities are
91 distributed mainly along a depth gradient (from the coastal zone to the upper slope) and
92 also along a longitudinal gradient, in associations with benthic macrofauna and

93 substratum (Gaertner et al. 1999, 2002, Gaertner 2000). These studies, however,
94 employed multivariate techniques, which do not include species distribution modeling.
95 Thus the present study investigated the habitat of ten demersal species in the Gulf of
96 Lions and predicted their spatial distribution in abundance using a GLM approach
97 incorporating four environmental variables (temperature, sediment, light penetration and
98 benthos macrofauna) as explanatory factors (McCullagh & Nelder 1989). The inter
99 annual stability of the spatial distributions of the ten species in May and June (Morfin et
100 al., 2012) makes it relevant to investigate their temporally persistent habitats.

101

102 From predicted maps of species densities (individuals per km²), we examined the
103 possibility to define an MPA for the ten species, using a MARXAN procedure (Ball &
104 Possingham 2000). MARXAN is one of the most widely used software for conservation
105 planning. It is a stochastic optimization procedure that identifies from a set of candidate
106 sites a subset of sites which achieves some particular biodiversity features, while
107 attempting to minimize the cost for their implementation (McDonnell et al. 2002). This
108 cost can be related to social, economical or ecological features, or a combination
109 thereof.

110

111 **Materials & Methods**

112 **Species data**

113 Species data were collected during the International Bottom Trawl Survey in the
114 Mediterranean Sea (MEDITS) conducted every year from 1994 to 2010, in May and
115 June. For the Gulf of Lions, 65 fixed sampling stations were chosen, using a random
116 stratified design and are sampled since 1994 (see Figure 1 and Bertrand et al. (2002)).
117 Trawling was operated during 30 minutes (until 200 m depth) and one hour (from 200 m
118 depth) in daylight following a standardized protocol (Fiorentini et al. 1999).
119 Catchability and accessibility were supposed to be constant over the whole area, though
120 the sites located on the slope were known to be less accessible to the trawl while adults
121 of several species, like *Merluccius merluccius*, are more abundant in deeper area. To
122 counterbalance this lower accessibility of the trawl, the MEDITS Group set a double
123 sampling duration in deeper waters (> 200 m, Bertrand et al. (2002) for more details).
124 Each catch content were then sorted by species, counted and weighted.

125 Since the onset of the MEDITS survey, 300 different species have been identified in the
126 Gulf of Lions. However, many are rare or low in abundance, which makes a habitat
127 model irrelevant for them (Mérigot et al. 2007). Therefore, we have only considered the
128 species present in at least 20% of the hauls and properly selected by the fishing gear.
129 Furthermore, we focused on exploited species and selected species at different trophic
130 and taxonomic (decapods, cephalopods, elasmobranchs, fish) positions (Bănaru et al.
131 2012, Morfin et al. 2012). This selection led us to retain the following ten species:
132 European hake (*Merluccius merluccius*), Atlantic horse mackerel (*Trachurus trachurus*),
133 Mediterranean horse mackerel (*Trachurus mediterraneus*), grey gurnard (*Eutrigla*
134 *gurnardus*), red gurnard (*Aspitrigla cuculus*), Norway lobster (*Nephrops norvegicus*),
135 red mullet (*Mullus barbatus*), horned octopus (*Eledone cirrhosa*), small-spotted
136 catshark (*Scyliorhinus canicula*) and elegant cuttlefish (*Sepia elegans*). Although both
137 *Trachurus* species are pelagic, we considered that they were properly sampled by the
138 MEDITS gear, as their distributions matched with those observed during the pelagic
139 acoustic survey PELMED, carried out in the same region.

140

141 **Environmental variables**

142 According to previous studies on fish habitat and data availability, we considered the
143 following five environmental variables: depth, biological zones, bottom temperature,
144 sediment type and benthos type (Johnson et al. 2012).

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

157 We thus considered biological zones as an alternative to bathymetry. Biological zones
158 were modelled by the EuseaMap project and defined by the percentage of light
159 penetration in the bottom, as a function of bathymetry and turbidity (Cameron & Askew
160 2011). The “infralittoral” zone was allocated to values up to 1% of light penetration, the
161 “circalittoral” zone to values between 1 and 5%, and the “bathyal” zone to values
162 greater than 5%. These thresholds (1% and 5%) are known to impact photosynthetic
163 activity and the presence of different kinds of algae that induce different biotopes
164 (Cameron & Askew 2011). From a statistical point of view, the relationships with
165 species distributions will be easier to establish as the coverage of the different zones are
166 balanced in the study area.

167 Bottom temperature was measured *in situ* during the survey by an onboard SCANMAR
168 device from 1996 to 2010. We used a variance decomposition method (Empirical
169 Orthogonal Functions) to demonstrate the high inter annual temporal stability of bottom
170 temperature the averaged values over the time period represented 78.2% of the total
171 variance of the bottom temperature (details in supplementary material S1). These results
172 legitimate the use in this study of averaged values of the bottom temperature over the
173 whole period, at each site.

174 A seabed sediment map was obtained from the EuSeamap project, which collated
175 sediment data from various sources and built a map at 250 m resolution, with four
176 sediment types: “gravels”, “sand”, “muddy-sand”, “sandy-mud”, and “mud” (Cameron
177 & Askew 2011). As only one site is associated to the “sand” class, gravels and sand
178 classes were included in the same “coarse sand” class for statistical purposes.

179 Benthos macrofauna is an important factor for the habitat of demersal fishes (see e.g
180 (Gaertner et al. 1999). Data on benthic species were collected by MEDITS survey, in
181 2000-2002, 2004, and 2006, and grouped into 15 categories. To reduce the number of
182 classes, we performed a cluster analysis, which led to retain three main benthos groups.
183 The three groups were dominated by Mantis shrimp and tunicates, and mostly differed
184 by their overall abundance whatever the type of fauna, rather than by their species
185 composition. The group 1 is the more abundant while the group 3 the less abundant.
186 Nonetheless, group 2 is the more abundant in sea urchins and polychaetes and group 3 is
187 the more abundant in Cnidaria, crabs and crinoids (Figure S2).

188

189

190 **Model formulation and calibration**

191 Relationships between species count data and environmental variables were analysed
192 following a GLM approach. In the present analysis, the retained variables of interest
193 were the number of individuals per species and site summed over the 17 years, as
194 Morfin et al. (2012) have shown that the spatial distribution of those species studied
195 were highly stable over time. The corresponding trawled surface (in km²) was accounted
196 in the model as an offset. For count data, Poisson and Negative Binomial (NB)
197 distributions are natural choices. Still, for all species the Poisson distribution appeared
198 to be clearly inappropriate according to the inspection of the residuals. We thus
199 developed NB models with a logarithm link function to approach linearity and maintain
200 model predictions within a range of values consistent with the original data (Guisan &
201 Zimmermann 2000). Nonetheless, non-linearities were detected between the log
202 expectancy of the response variable and the bottom temperature (the only continuous
203 covariate). The bottom temperature (BT) was thus tested using a first, second and third
204 order polynomial terms, to reflect a potential optimum (second degree) in the response
205 of species abundance to temperature, and a skewed form of the response around the
206 optimum (third degree). The interactions between temperature and biological zones
207 were also tested, resulting in 92 candidate models.

208 As interactions and polynomial terms are often the source of multicollinearities, one
209 possible multicollinearity diagnostic is the variance inflation factor (VIF) analysis,
210 which is generalized for categorical variables in GVIF (Montgomery & Peck 1982). As
211 a rule of thumb, multicollinearity may be problematic (increase the parameters
212 variances and arise interpretation difficulties) when the $GVIF^{1/2df}$ is greater than 2 (Fox
213 & Monette 1992). This problem was solved for polynomial terms by orthogonalizing
214 them, but multicollinearities were still found for interaction terms.

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

218

219 **Model selection**

220 Two criteria were used to select models, a corrected Akaike Information Criterion
221 (AICc) and prediction performance index estimated by cross-validation. The necessary
222 bias correction of AIC for small sample size was estimated by Gaussian approximation.

223 In addition, AIC-based selection may result in overfitted models and/or in inappropriate
224 promotion of complex hypotheses (Burnham & Anderson 2003). Cross-validation
225 procedures have been proposed to overcome this problem and estimate how accurately a
226 predictive model will perform in practice (Geisser 1975, Efron & Tibshirani 1993).
227 Though less robust than “k-fold” cross-validation procedure, leave-one-out (Loo) cross-
228 validation was chosen to keep enough data in the “training” sample (Arlot & Celisse
229 2010). Prediction performance was then measured by the Percentage of Mean Absolute
230 Error (PMAE), i.e. MAE divided by mean abundance, on the data left out. A value
231 greater than one indicates that the average prediction error is higher than the average
232 abundance. Both selection procedures, AICc and Loo, were applied for comparison
233 purposes, but we used models selected by Loo for the species distribution predictions
234 (see below) as the criterion is based on the prediction performance.

235

236 **Model evaluation**

237 Species distribution models were evaluated using residual analysis and deviance
238 estimates. A quantile-quantile plot of standardized residuals, deviance, was used to
239 check the assumption on the model distribution residuals. Furthermore plots of these
240 residuals against fitted values and each explanatory variable allowed for identifying
241 unexpected patterns in the deviance. The proportion of deviance explained by the
242 predictors was also calculated to assess the explanatory power of the model.

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

248

249 **Species Distribution Maps**

250 Predictions of species abundance were calculated from habitat models selected by the
251 Loo procedure, as linear combinations of the explanatory variables on a 2' x 2' grid
252 (Table 1). Species distributions over the Gulf of Lions were thus built using maps for
253 each of the four predictor variables to predict each species' habitat (Figure 2). The maps
254 of benthos groups and bottom temperature were not available over the whole study area

255 and were thus interpolated. Bottom temperature was interpolated on the prediction grid
256 by ordinary kriging (Matheron 1963, Cressie 1993). As the benthos is a categorical
257 variable, values were predicted using the Voronoï polygons (each pixel was associated
258 to the group of the nearest observed site).

259

260 **MARXAN analysis**

261 MARXAN is an optimisation algorithm which implements an objective function to
262 minimize, including a penalization term for not achieving the conservation target and
263 the cost of the reserve. The planning units were created by dividing the study area into
264 2'x2' squares and the target objectives were species distribution maps predicted by
265 GLMs. The biodiversity target was formulated to ensure that at least 20% of abundance
266 of each species was represented in a protected area network. This threshold was
267 proposed at the 2002 Earth Summit and advocated as the minimum amount of each
268 habitat to be represented in marine reserves (IUCN World Parks Congress 2003).
269 Reserve area was used as a surrogate for cost the planning units, based on the
270 assumption that the larger the reserve, the more costly the implementation and the
271 management. As the optimal solution may be highly fragmented, a penalty for the total
272 boundary length of the reserve (boundary length modifier, BLM) was included to get
273 the best compromise between the total area of the conservation system and its
274 compactness (Stewart & Possingham 2005).

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

285

286

287 **Results**

288 **Model evaluation and interpretation**

289 For all species, graphics of deviance residuals *versus* fitted values did not display any
290 special pattern and less than 5% of the values were found outside the 95%-confidence
291 interval (Figure S3). Quantile-quantile plots displayed no significant departure of
292 deviance residuals from normal distribution, except for red gurnard (Figure S4).

293 Deviations at extremities were observed for several species, still the normality of error
294 is not a condition of GLM quality but simply a description of model behaviour.

295 Variograms of deviance residuals presented some auto-correlation structure for three
296 species: hake, catshark and octopus (Figure S5). For the two mackerels, the spatial
297 structure is only present with the model selected by Loo.

298 Several differences were detected between models selected with the Loo and AICc.
299 However, no general rule can be established, except that for the same level of
300 complexity the factors selected by AICc explained slightly higher percentages of
301 deviance (in average 67.4% and 70.9% respectively, Table 1). Overall the percentages
302 of explained deviance were substantial, ranging from 31.6% (for red gurnard) to 89.6%
303 (for Norway lobster) and correlated to number of parameters, varying from 6 to 15.

304 Although no significant linear correlation was detected between covariates, there were
305 some redundancies that make not possible to distinguish the part of deviance explained
306 by each covariate. That is why the sum of deviance explained by each covariate
307 separately may be much greater than the total explained deviance, e.g. Norway lobster,
308 strongly associated with the two biological zones and the bottom temperature (Table 2).

309 However, these results highlight that the factors that were not selected by both selection
310 procedures generally explained very low percentage of deviance, except in two cases.

311 Depending on the model selection, a strong association is established between red
312 gurnard distribution and the temperature or substrate type; and between Norway lobster
313 and benthos type and substratum. For a given species, the factors selected by both
314 procedures displayed the same marginal effects. We thus focused on these factors for
315 model interpretation.

316 The biological zone variable was systematically selected except for small-spotted
317 catshark and red gurnard. This factor explained 31.2% of species distribution deviance
318 on average, with a maximum of 80.8% for Norway lobster. Cuttlefish was strongly

319 associated with the circalittoral zone while Norway lobster with the bathyal zone. The
320 seven other species were associated with both circalittoral and infralittoral zones
321 (Figure S6).

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

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

335 Benthos groups explained 6.7% of the deviance in average, with a maximum of 50.7%
336 for Norway lobster. The group 2 was the referential because species were ordered in
337 abundance, though each group displayed some differences in species composition (see
338 Material and Methods). Red mullet, red gurnard and catshark were associated to low
339 benthos abundance level, grey gurnard and Atlantic horse mackerel to intermediate
340 level, and hake and Norway lobster to high level (Figure S8).

341 The predictive power of models was measured by PMAE, a value upper than one
342 indicating a poor predictive capacity. Except for red gurnard (PMAE=114.6%), for the
343 nine other species models selected by Loo had satisfying prediction performance
344 (PMAE=30-75%, Table 1). The species distributions maps were thus predicted using the
345 models selected by Loo.

346

347 **Species distribution maps**

348 Mapping model predictions over the Gulf of Lions highlighted, as expected, different
349 spatial patterns across species (Figure 4). Species found close to the coasts were horse
350 mackerels, cuttlefish and red mullet. Atlantic horse mackerel was also found in the west

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

356 Prediction uncertainty was measured by the coefficient of variation (CV) of predictions,
357 a value greater than 1 indicating a large uncertainty in prediction. CV were generally
358 lower than 1, with median values ranging from 0.16 (for horned octopus) to 0.71 (for
359 Norway lobster), reflecting overall good predictions. Furthermore, values of CV were
360 consistent with the index of prediction performance, PMAE, as median values of CV
361 decreased with increasing PMAE (Table 1). Poorest predictions ($CV > 1$) appeared due to
362 values of the explanatory variables outside the range used to calibrate the models. These
363 values covered 2% of the ten maps and were mostly detected in the eastern side of the
364 slope and in the western side of the shelf, where temperature was not directly measured,
365 but interpolated by kriging. In those areas, kriged values were significantly higher than
366 the observations. The predictions for which CV were greater than one were discarded
367 for the MARXAN analysis.

368

369 **MARXAN analysis**

370 Among 500 runs using different initializing values, we looked at the best solution and
371 the frequency of selection of each planning unit (note that the penalty for not achieving
372 the conservation targets in the objective function was set to the minimum value where
373 all conservation targets were met). Without any boundary length constraint the set of
374 planning units selected was highly scattered over the whole area, reflecting both the
375 sparsity of some species distributions, e.g. grey gurnard, and the differences between
376 species distributions, e.g. Norway lobster *versus* cuttlefish, the former being offshore
377 while the latter is coastal (Figure 5).

378 The weight allocated to boundary length penalty (named Boundary Length Modifier
379 BLM), was set in order to get the best trade-off between the boundary length and the
380 total area of the reserve system (Figure S10). The more optimal planning design
381 solution according to the BLM (0.25) occupied 16.4% of the study region and was
382 composed of seven disjointed zones (Figure 4, second line of the right column). This

383 solution includes the 3.2% of planning units selected more than 80% of times and 67%
384 was selected more than 50% of times. Among the seven zones, one located on the west
385 coast is a preferential habitat for the two mackerels, grey gurnard, horned octopus and
386 cuttlefish, while on the middle of the coast only both mackerels and cuttlefish were
387 found. Just next, three tiny coastal areas were selected because red gurnard is associated
388 to their sandy bottoms (Figure 4). The last zone on the coast is the eastern area where
389 hake and horned octopus were abundant. The two zones in the middle of the shelf
390 corresponding to sandy bottoms patches were particularly important as all species
391 except the two mackerels and cuttlefish are found in abundance (Figure 2). The last
392 zone located on the slope was the only area where the most preferential habitat of
393 Norway lobster and some other species overlapped. Furthermore they included the main
394 types of physical habitats present in the Gulf of Lions in similar proportions, i.e. all the
395 biological zones, the full range of bottom temperature, all the sediment types and the
396 three benthos groups.

397

398 **Discussion**

399 In this study we characterized habitats and predict the distributions for ten demersal
400 species in the Gulf of Lions, using a GLM approach. The selected species distribution
401 models have a good explanatory power as the percentages of explained deviance were
402 high for all species. Their predictive capabilities were also satisfying, except for red
403 gurnard. For this species, the model performance was probably affected by the fact that
404 its highest abundance was observed in the only sandy site. This strong association could
405 not be clearly detected as we had no choice to mix this site with gravels bottom sites in
406 the same “coarse-sand” category, for statistical significance.

407

408 We compared the results from two different model selection procedures, Loo cross-
409 validation and AIC. Overall, results were quite similar. In theory, the former should be
410 preferred for predictive models while the latter should be preferred for explanatory
411 models. Still, there is no rule of thumb to choose a model selection procedure, and our
412 goal was to highlight the uncertainty that it may involve. Unexpectedly, the AIC based
413 procedure did not systematically select models with a higher explanatory power than the
414 cross-validation procedure. While we tested several degrees of polynomials for

415 temperature, we kept fixed the number of classes for categorical variables. Thus, models
416 including four classes of sediment type were systematically more penalized than the
417 others. That is why any interpretation about identified habitat factors and the
418 exploitation of predicted maps should be taken with caution if these aspects were
419 investigated.

420

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

434

435 Searching for significant relationships between species abundance and measured habitat
436 variables has been criticized for ascribing coincidental correlations or indirect
437 relationships as direct causal links (Guisan & Thuiller 2005). However, insufficient
438 knowledge still remain on factors influencing marine species, which make difficult to
439 establish prior assumptions about causal relationships and to test them. Consequently,
440 correlative approaches that make few or even no prior assumptions about underlying
441 causal relationships are considered legitimate when attempting to understand the
442 complex interactions between fish populations and their environment (Valavanis et al.
443 2004). Those correlations can in turn be used as the basis for subsequent hypothesis-
444 driven studies aiming at determining demersal fish habitat requirements. However,
445 model outputs must be interpreted with caution when the sampling design could not be
446 totally controlled and balanced. For instance the range 13-13.5°C was only observed in

447 the bathyal zone, where the relationship between temperature and the log of the species
448 abundance was positive for several species, while it was negative in other ranges. In this
449 case, interactions may be only statistics and reflect the non-linearity of the relationship
450 between temperature and the log of the response variable, instead of a variation in the
451 effect of temperature depending on biological zones.

452

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

463

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

480 bathymetry as the distribution of the latter is highly skewed in the Gulf of Lions, which
481 made difficult to model it and led to less satisfactory models (according to both model
482 selection criteria) than the models using the Biological Zones.

483

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

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

500

501 Distributions of species studied here were highly stable over the whole period 1994-
502 2010 (Morfin et al. 2012). This makes relevant the present study goal to determine
503 determining species habitat that persisted over time. The strong association
504 demonstrated in this analysis between averaged species abundance and temporally
505 rather stable external factors (substrate, bottom temperature and biological zones) would
506 certainly be still important in habitat models including temporal variability. Despite a
507 strong temporal stability, Morfin et al. (2012) also documented a positive relationship
508 between species occupancy area and total abundance over the 17 years. According to
509 MacCall basin theory, habitat selection is density-dependent to some degree. At low
510 abundance, individuals occupy the most suitable habitat, whereas they are expected to
511 spread to marginal (sub-optimal) habitats when abundance increases because of

512 intraspecific competition for food and/or space (Fretwell & Lucas 1970, MacCall 1990).
513 The results of the present paper demonstrate that spatial distribution of species are
514 strongly associated to environmental factors. Furthermore, these associations were
515 established from data averaged over the time period, and are thus independent from
516 temporally variable habitats. Therefore, it is more likely that a spatial expansion of
517 species at high abundance is due to a density-dependent process, as expected under
518 MacCall basin theory, rather than to spatio-temporal variations in some key
519 environmental variables. Therefore, we can expect that the deviance explained by these
520 factors in temporal models would be lower if the density-dependent process is not taken
521 into account.

522

523 We performed a first quantitative analysis to investigate the potential relevance of an
524 MPA in the Gulf of Lions, using MARXAN procedure (Ball & Possingham 2000). Our
525 goal was to determine the minimum set of areas containing at least 20% of abundance
526 of each of the ten target species, which is obviously more complicated than for one
527 given species (Fromentin & Lopuszanski 2013). The first solution obtained without any
528 boundary length constraint, was a set of areas highly scattered over the whole Gulf of
529 Lions, reflecting the sparsity of some species distributions as well as the differences
530 between species distributions. Such a solution was obviously not operational in terms of
531 management. Once increasing the compactness of the protected areas by increasing its
532 boundary length, we obtained more “manageable” solutions. The design realising the
533 best compromise between total area and compactness covered 15.6% of the study region
534 and was comprised of seven disjointed zones. As this design was still very sparse, we
535 also considered a higher compactness constraint, which produced three distinct zones
536 covering 17% of the study area (Figure 4, third line of the right column). This
537 percentage should be interpreted relatively to the minimum area of 12% that 20% of
538 horned octopus distribution covers. The high coverage appears more due to the sparsity
539 of this species than the variability between species distributions. The conservation
540 objective should probably be lowered as this region is highly exploited and closing 20%
541 of the fishing grounds would probably be regarded as unacceptable by some
542 stakeholders. Some choices and compromises at the ecological/scientific level, but also
543 at the economic and political levels would be required (Sumaila et al. 2007, Yates &
544 Schoeman 2013). Although this is beyond the scope of the present study, assessing the

545 potential benefits of a MPA should also consider the effects of the redistribution of
546 fishing effort outside the MPA and the spatial dynamics of species of the main species
547 (Apostolaki et al. 2002).

548 As the surveys are only carried out in May-June, this study can hardly tackle seasonal
549 issues. This period has been selected as it is a recruitment period for many species. This
550 might be problematic, as all species, excepting Norway lobster, make ontogenic
551 migrations, which may induce seasonal variability. For many species, juveniles migrate
552 between the coastal and bathyal zones according to seasons, as nurseries are located in
553 both areas (GFCM 2010). Adults of hake, the most documented species, migrate to the
554 middle of the shelf to spawn all over the year, the spawning peak being in winter and
555 early spring. Nevertheless, too few information are available on species distributions in
556 the Gulf of Lions during the other periods of the year. However, it would be interesting
557 to perform the same analysis on juveniles and adults separately to identify nurseries and
558 spawning areas. The potential habitat of young-of-the year individuals of hake
559 determined by Druon et al. (2015) are consistent with our results (temperature range and
560 bottom type). Colloca et al. (2015) investigated the overlap of several demersal species
561 nurseries, using spatial models. Although the overlap covered a too large area according
562 to fishing activity, such an approach deserves further attention. The study of adults
563 distributions requires models fitted to zero-inflated data, as adult specimens are
564 observed in a very few proportion of sampled sites (Heilbron 1994).

565

566 Finally this analysis indicates that designing an MPA network for several species of
567 interest in the Gulf of Lions is not straightforward and deserves more dedicated
568 investigations and co-constructions processes. Furthermore, several sources of
569 uncertainties were highlighted throughout the analysis (model selection, predictions,
570 optimization algorithm), which could lead to spurious conclusions. Bayesian framework
571 is well adapted for uncertainty propagation, and new advances to integrate map
572 uncertainty in MPA design software are on their way (Carvalho et al. 2011, Kujala et al.
573 2013).

574

575

576 **Supplementary materials**

577 S1: Empirical Orthogonal Functions analysis on bottom temperature.

578 S2: Benthos groups.

579 S3: Deviance residuals versus the fitted values.

580 S4: Quantile-quantile plots of deviance residuals.

581 S5: Variograms of deviance residuals.

582 S6-S8: Marginal effects of biological zones, substrate and benthos groups.

583 S9: Species density maps.

584 S10: Boundary Length Modifier analysis.

585

586 **Acknowledgments**

587 Thanks are due to Ifremer-Sète laboratory, which has carried out the MEDITS trawl
588 survey in the Gulf of Lions every year since 1994. MEDITS trawl survey is funded by
589 the Directorate-General for Maritime Affairs and Fisheries (DG-MARE) (European
590 Commission) and Ifremer. We are also grateful to Ifremer-Brest, which provided us
591 advices and data from Euseamap project, to Tristan Rouyer, who improved the English
592 quality of the paper and to the three anonymous reviewers for their very substantial
593 work. MM was supported by the AMPED project (www.amped.ird.fr) through a grant
594 from the French National Research Agency (ANR), Systerra Program, grant number
595 ANR-08-STRA-03.

596

597 **References**

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

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

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

605 Ball IR, Possingham HP (2000) MARXAN (V1. 8.2) User's Manual.

- 606 Bănaru D, Mellon-Duval C, Roos D, Bigot J-L, Souplet A, Jadaud A, Beaubrun P,
607 Fromentin J-M (2012) Trophic structure in the Gulf of Lions marine ecosystem
608 (north-western Mediterranean Sea) and fishing impacts. *J Mar Syst*
- 609 Bartolino V, Ottavi A, Colloca F, Ardizzone GD, Stefánsson G (2008) Bathymetric
610 preferences of juvenile European hake (*Merluccius merluccius*). *ICES J Mar Sci*
611 *J Cons* 65:963–969
- 612 Bertrand JA, Gil De Sola, Luis, Papaconstantinou C, Relini G, Souplet A (2002) The
613 general specifications of the MEDITS surveys. *Sci Mar* 66:9–17
- 614 Bethoux JP, Gentili B, Raunet J, Tailliez D (1990) Warming trend in the western
615 Mediterranean deep water. *Nature* 347:660–662
- 616 Bianchi CN, Morri C (2000) Marine Biodiversity of the Mediterranean Sea: Situation,
617 Problems and Prospects for Future Research. *Mar Pollut Bull* 40:367–376
- 618 Breiman L (2001) Random Forests. *Mach Learn* 45:5–32
- 619 Burnham KP, Anderson DR (2003) *Model Selection and Multi-Model Inference: A*
620 *Practical Information-Theoretic Approach*, 2nd ed. 2002. Corr. 3rd printing.
621 Springer
- 622 Caddy JF, Refk R, Do-Chi T (1995) Productivity estimates for the Mediterranean:
623 evidence of accelerating ecological change. *Ocean Coast Manag* 26:1–18
- 624 Cameron A, Askew N (2011) *EUSeaMap-Preparatory Action for development and*
625 *assessment of a European broad-scale seabed habitat map final report*. Cameron
626 Askew NEds:240
- 627 Carlucci R, Giuseppe L, Porzia M, Francesca C, Alessandra MC, Letizia S, Teresa SM,
628 Nicola U, Angelo T, D’Onghia G (2009) Nursery areas of red mullet (*Mullus*
629 *barbatus*), hake (*Merluccius merluccius*) and deep-water rose shrimp
630 (*Parapenaeus longirostris*) in the Eastern-Central Mediterranean Sea. *Estuar*
631 *Coast Shelf Sci* 83:529–538
- 632 Carvalho SB, Brito JC, Crespo EG, Watts ME, Possingham HP (2011) Conservation
633 planning under climate change: Toward accounting for uncertainty in predicted
634 species distributions to increase confidence in conservation investments in space
635 and time. *Biol Conserv* 144:2020–2030
- 636 Colloca F, Garofalo G, Bitetto I, Facchini MT, Grati F, Martiradonna A, Mastrantonio G,
637 Nikolioudakis N, Ordinas F, Scarcella G, Tserpes G, Tugores MP, Valavanis V,
638 Carlucci R, Fiorentino F, Follesa MC, Iglesias M, Knittweis L, Lefkaditou E,
639 Lembo G, Manfredi C, Massuti E, Pace ML, Papadopoulou N, Sartor P, Smith
640 CJ, Spedicato MT (2015) The Seascape of Demersal Fish Nursery Areas in the
641 North Mediterranean Sea, a First Step Towards the Implementation of Spatial
642 Planning for Trawl Fisheries. *PLoS ONE* 10:e0119590
- 643 Coll M, Piroddi C, Albouy C, Ben Rais Lasram F, Cheung WWL, Christensen V,
644 Karpouzi VS, Guilhaumon F, Mouillot D, Paleczny M, Palomares ML,
645 Steenbeek J, Trujillo P, Watson R, Pauly D (2012) The Mediterranean Sea under

- 646 siege: spatial overlap between marine biodiversity, cumulative threats and
647 marine reserves. *Glob Ecol Biogeogr* 21:465–480
- 648 Coll M, Piroddi C, Steenbeek J, Kaschner K, Ben Rais Lasram F, Aguzzi J, Ballesteros
649 E, Bianchi CN, Corbera J, Dailianis T, Danovaro R, Estrada M, Frogia C, Galil
650 BS, Gasol JM, Gertwagen R, Gil J, Guilhaumon F, Kesner-Reyes K, Kitsos M-S,
651 Koukouras A, Lampadariou N, Laxamana E, López-Fé de la Cuadra CM, Lotze
652 HK, Martin D, Mouillot D, Oro D, Raicevich S, Rius-Barile J, Saiz-Salinas JI,
653 San Vicente C, Somot S, Templado J, Turon X, Vafidis D, Villanueva R,
654 Voultziadou E (2010) The Biodiversity of the Mediterranean Sea: Estimates,
655 Patterns, and Threats. *PLoS ONE* 5:e11842
- 656 Cressie N (1993) *Statistics for Spatial Data, Revised Edition*. Wiley-Interscience, New
657 York
- 658 Druon J-N, Fiorentino F, Murenu M, Knittweis L, Colloca F, Osio C, Mérigot B,
659 Garofalo G, Mannini A, Jadaud A, Sbrana M, Scarcella G, Tserpes G, Peristeraki
660 P, Carlucci R, Heikkonen J (2015) Modelling of European hake nurseries in the
661 Mediterranean Sea: An ecological niche approach. *Prog Oceanogr* 130:188–204
- 662 Efron B, Tibshirani R (1993) *An introduction to the bootstrap*. CRC Press
- 663 Elith J, Leathwick JR (2009) *Species Distribution Models: Ecological Explanation and*
664 *Prediction Across Space and Time*. *Annu Rev Ecol Evol Syst* 40:677–697
- 665 Ferraro SP, Cole FA (2007) Benthic macrofauna–habitat associations in Willapa Bay,
666 Washington, USA. *Estuar Coast Shelf Sci* 71:491–507
- 667 Fiorentini L, Dremière LY, Leonori I, Sala A, Palumbo V (1999) Efficiency of the
668 bottom trawl used for the Mediterranean international trawl survey (MEDITS).
669 *Aquat Living Resour* 12:187–205
- 670 Fox J, Monette G (1992) Generalized collinearity diagnostics. *J Am Stat Assoc* 87:178–
671 183
- 672 Freire J, Garcia-Allut A (2000) Socioeconomic and biological causes of management
673 failures in European artisanal fisheries: the case of Galicia (NW Spain). *Mar*
674 *Policy* 24:375–384
- 675 Fretwell S, Lucas H (1970) On territorial behaviour and other factors influencing habitat
676 distribution in birds. *Acta Biotheor* 19:16–36
- 677 Fromentin J-M, Lopuszanski D (2013) Migration, residency, and homing of bluefin tuna
678 in the western Mediterranean Sea. *ICES J Mar Sci J Cons*:fst157
- 679 Gaertner J-C (2000) Seasonal Organization Patterns of Demersal Assemblages in the
680 Gulf of Lions (north-Western Mediterranean Sea). *J Mar Biol Assoc U K*
681 80:777–783
- 682 Gaertner J-C, Bertrand JA, Relini G, Papaconstantinou C, Mazouni N, Sola LG de,
683 Durbec J, JukicPeladic S, Souplet A (2007) Spatial pattern in species richness of

- 684 demersal fish assemblages on the continental shelf of the northern
685 Mediterranean Sea: a multiscale analysis. *Mar Ecol Prog Ser* 341:191–203
- 686 Gaertner J-C, Bertrand JA, Souplet A (2002) STATIS-CoA: A methodological solution
687 to assess the spatio-temporal organization of species assemblages. Application to
688 the demersal assemblages of the French Mediterranean Sea. *Sci Mar* 66:221–232
- 689 Gaertner JC, Mazouni N, Sabatier R, Millet B (1999) Spatial structure and habitat
690 associations of demersal assemblages in the Gulf of Lions: a
691 multicompartmental approach. *Mar Biol* 135:199–208
- 692 Geisser S (1975) The predictive sample reuse method with applications. *J Am Stat*
693 *Assoc* 70:320–328
- 694 GFCM (2010) Report of the twelfth session of the Scientific Advisory Committee. FAO
695 Fisheries and Aquaculture, Budva, Montenegro
- 696 GFCM (2011) Report of the thirteenth session of the Scientific Advisory Committee.
697 FAO Fisheries and Aquaculture, Marseille, France
- 698 Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple
699 habitat models. *Ecol Lett* 8:993–1009
- 700 Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology.
701 *Ecol Model* 135:147–186
- 702 Heilbron DC (1994) Zero-altered and other regression models for count data with added
703 zeros. *Biom J* 36:531–547
- 704 IUCN World Parks Congress (2003) Vth IUCN World Parks Congress. IUCN, Durban,
705 South Africa
- 706 Johnson AF, Jenkins SR, Hiddink JG, Hinz H (2012) Linking temperate demersal fish
707 species to habitat: scales, patterns and future directions. *Fish Fish*
- 708 Kaiser MJ, Rogers SI, Ellis JR (1999) Importance of benthic habitat complexity for
709 demersal fish assemblages. In: *American Fisheries Society Symposium*.p 212–
710 223
- 711 Kujala H, Moilanen A, Araújo MB, Cabeza M (2013) Conservation Planning with
712 Uncertain Climate Change Projections. *PLoS ONE* 8:e53315
- 713 MacCall AD (1990) *Dynamic Geography of Marine Fish Populations*. Univ of
714 Washington Pr, Seattle
- 715 Matheron G (1963) Principles of geostatistics. *Econ Geol* 58:1246–1266
- 716 McCullagh P, Nelder JA (1989) *Generalized Linear Models*, Second Edition. CRC Press
- 717 McCullough BD (1994) Bootstrapping forecast intervals: An application to AR(p)
718 models. *J Forecast* 13:51–66

- 719 McDonnell MD, Possingham HP, Ball IR, Cousins EA (2002) Mathematical methods
720 for spatially cohesive reserve design. *Environ Model Assess* 7:107–114
- 721 Mérigot B, Bertrand JA, Mazouni N, Manté C, Durbec J-P, Gaertner J-C (2007) A
722 multi-component analysis of species diversity of groundfish assemblages on the
723 continental shelf of the Gulf of Lions (north-western Mediterranean Sea). *Estuar
724 Coast Shelf Sci* 73:123–136
- 725 Millot C (1990) The Gulf of Lions' hydrodynamics. *Cont Shelf Res* 10:885–894
- 726 Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux
727 J, Fonseca GAB da (2005) Hotspots Revisited: Earth's Biologically Richest and
728 Most Endangered Terrestrial Ecoregions. Conservation International,
729 Whashington D.C.
- 730 Montgomery DC, Peck EA (1982) Introduction to linear regression analysis. Wiley
- 731 Morfin M, Fromentin J-M, Jadaud A, Bez N (2012) Spatio-Temporal Patterns of Key
732 Exploited Marine Species in the Northwestern Mediterranean Sea. *PLoS ONE*
733 7:e37907
- 734 Murawski SA, Finn JT (1988) Biological bases for mixed-species fisheries: species co-
735 distribution in relation to environmental and biotic variables. *Can J Fish Aquat
736 Sci* 45:1720–1735
- 737 Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB da, Kent J (2000)
738 Biodiversity hotspots for conservation priorities. *Nature* 403:853–858
- 739 Nicolas D, Le Loc'h F, Desaunay Y, Hamon D, Blanchet A, Le Pape O (2007)
740 Relationships between benthic macrofauna and habitat suitability for juvenile
741 common sole (*Solea solea*, L.) in the Vilaine estuary (Bay of Biscay,
742 France) nursery ground. *Estuar Coast Shelf Sci* 73:639–650
- 743 Papaconstantinou C, Farrugio H (2000) Fisheries in the Mediterranean. *Mediterr Mar
744 Sci* 1:5–18
- 745 Persohn C, Lorance P, Trenkel VM (2009) Habitat preferences of selected demersal fish
746 species in the Bay of Biscay and Celtic Sea, North - East Atlantic. *Fish
747 Oceanogr* 18:268–285
- 748 Quéro J-C, Vayne J-J (1997) Les poissons de mer des pêches françaises. Delachaux &
749 Niestle
- 750 Ramírez-Bastida P, Navarro-Sigüenza AG, Peterson AT (2008) Aquatic bird distributions
751 in Mexico: designing conservation approaches quantitatively. *Biodivers Conserv*
752 17:2525–2558
- 753 R Development Core Team (2011) R: A Language and Environment for Statistical
754 Computing. R Foundation for Statistical Computing, Vienna, Austria
- 755 Roberts CM, Andelman S, Branch G, Bustamante RH, Carlos Castilla J, Dugan J,
756 Halpern BS, Lafferty KD, Leslie H, Lubchenco J, McArdle D, Possingham HP,

- 757 Ruckelshaus M, Warner RR (2003) Ecological criteria for evaluating candidate
758 sites for marine reserves. *Ecol Appl* 13:199–214
- 759 Sanchez-Vidal A, Pasqual C, Kerhervé P, Heussner S, Calafat A, Palanques A, Durrieu
760 de Madron X, Canals M, Puig P (2009) Across margin export of organic matter
761 by cascading events traced by stable isotopes, northwestern Mediterranean Sea.
762 *Limnol Ocean* 54:1488–1500
- 763 Stewart RR, Possingham HP (2005) Efficiency, costs and trade-offs in marine reserve
764 system design. *Environ Model Assess* 10:203–213
- 765 Sumaila U, Zeller D, Watson R, Alder J, Pauly D (2007) Potential costs and benefits of
766 marine reserves in the high seas. *Mar Ecol Prog Ser* 345:305–310
- 767 Valavanis VD, Georgakarakos S, Kapantagakis A, Palialexis A, Katara I (2004) A GIS
768 environmental modelling approach to essential fish habitat designation. *Ecol*
769 *Model* 178:417–427
- 770 Venables WN, Ripley BD (1994) *Modern applied statistics with S-PLUS*. Springer-
771 verlag New York
- 772 Yates KL, Schoeman DS (2013) Spatial Access Priority Mapping (SAPM) with Fishers:
773 A Quantitative GIS Method for Participatory Planning. *PLoS ONE* 8:e68424
- 774 Yoklavich M, Greene HG, Cailliet GM, Sullivan DE, Lea RN, Love MS (2000) Habitat
775 associations of deep-water rockfishes in a submarine canyon: an example of a
776 natural refuge. *Fish Bull* 98

777

778

779 **Tables**

780

781 **Table 1.** Summary of selected models for the ten demersal species, using binomial
782 negative regressions. For each species, first line includes results from the model selected
783 by Leave-one-out cross-validation procedure, and second line (in italic) includes results
784 from the model selected according to AICc. Predictors are biological zone (BZ), bottom
785 temperature (T), substrate (SUB), benthos (BENT) and slope (S). Symbol ‘*’ between
786 predictors indicates interactions between model predictors. “Ex Dev” column contains
787 percentages of deviance explained by the covariates, “BZ Ex Dev” contains percentages
788 of deviance explained by BZ variable only. “PMAE” is the Percentage of Mean
789 Absolute Error calculated by cross-validation, an indicator of model prediction power.

790

Species	Selected model (number of parameters)	Ex Dev (%)	PMAE (%)
Hake	ZB+T+T ² +BENT+ZB*(T+T ²) (12)	59.3	0.33
	ZB+T+SUB+ZB*T (10)	61.5	0.46
Atlantic horse mackerel	ZB+T+BENT (7)	46.5	0.58
	ZB+T+SUB+ZB*T (10)	76.3	0.63
Mediterranean horse mackerel	ZB+T+SUB (8)	48.8	0.69
	ZB+T+T ² +T ³ +ZB*(T+T ² +T ³) (13)	72.2	0.72
Grey gurnard	ZB+T+T ² +SUB+BENT+ZB*(T+T ²) (15)	86.5	0.40
	ZB+T+T ² +BENT+ZB*(T+T ²) (12)	85.2	0.40
Red gurnard	SUB+BENT (7)	31.6	11.46
	ZB+T+BENT+ZB*T (9)	54.0	1513.8
Norway lobster	ZB+T+T ² +BENT+ZB*(T+T ²) (12)	87.2	0.78
	ZB+T+T ³ +SUB+ZB*T ³ (12)	89.6	0.79
Red mullet	ZB+T+T ² +SUB+BENT+ZB*(T+T ²) (15)	81.1	0.64
	ZB+T ³ +SUB+BENT+ZB*(T ³) (12)	79.2	0.72
Octopus	ZB+T+T ² +SUB (9)	61.3	0.30
	ZB+T+T ² +ZB*(T+T ²) (9)	66.4	0.31
Small-spotted catshark	T+SUB+BENT (8)	46.7	0.75
	T+T ² +T ³ +SUB (8)	49.3	0.79
Cuttlefish	ZB+T+T ² +T ³ +BENT (9)	76.5	0.47
	ZB+T+T ² (6)	75.6	0.49

792 **Table 2.** Proportion of deviance explained by each covariate included in the selected
 793 model.

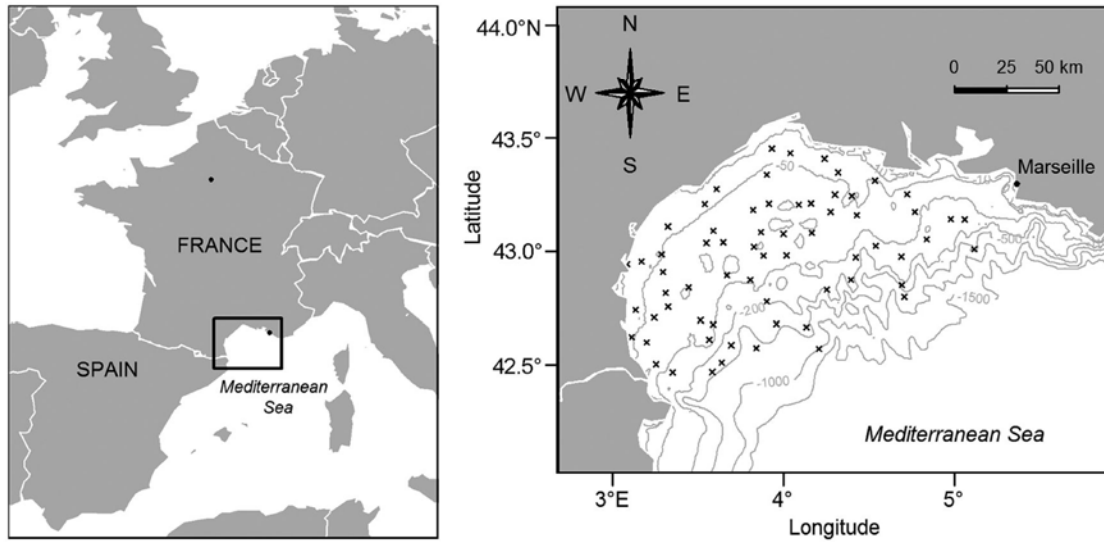
794

Species	Biological Zones	Temperature	Benthos	Substratum	Interaction BZ:T
Hake	20.7/20.9	2.7/2.0	1.8/NA	NA/9.4	32.0/25.7
Atlantic horse mackerel	42.0/43.0	10.2/10.4	3.4/NA	NA/3.7	NA/27.4
Mediterranean horse mackerel	33.9/35.0	18.6/34.6	NA	2.8/NA	NA/25.7
Grey gurnard	19.3/23.7	3.0/3.7	5.3/6.7	2.0/NA	55.7/56.0
Red gurnard	NA/5.0	NA/14.2	14.7/15.0	30.0/NA	NA/16.1
Norway lobster	70.8/71.6	56.2/63.0	50.7/NA	NA/39.8	8.2/6.0
Red mullet	14.8/14.7	1.3/0.3	11.6/11.7	42.8/43.1	10.1/22.5
Octopus	37.9/38.3	22.5/22.7	NA	4.7/NA	NA/6.4
Small-spotted catshark	NA	35.1/41.0	10.0/NA	8.8	NA
Cuttlefish	65.8	48.6/44.8	3.5/NA	NA	NA

795

796 **Figures**

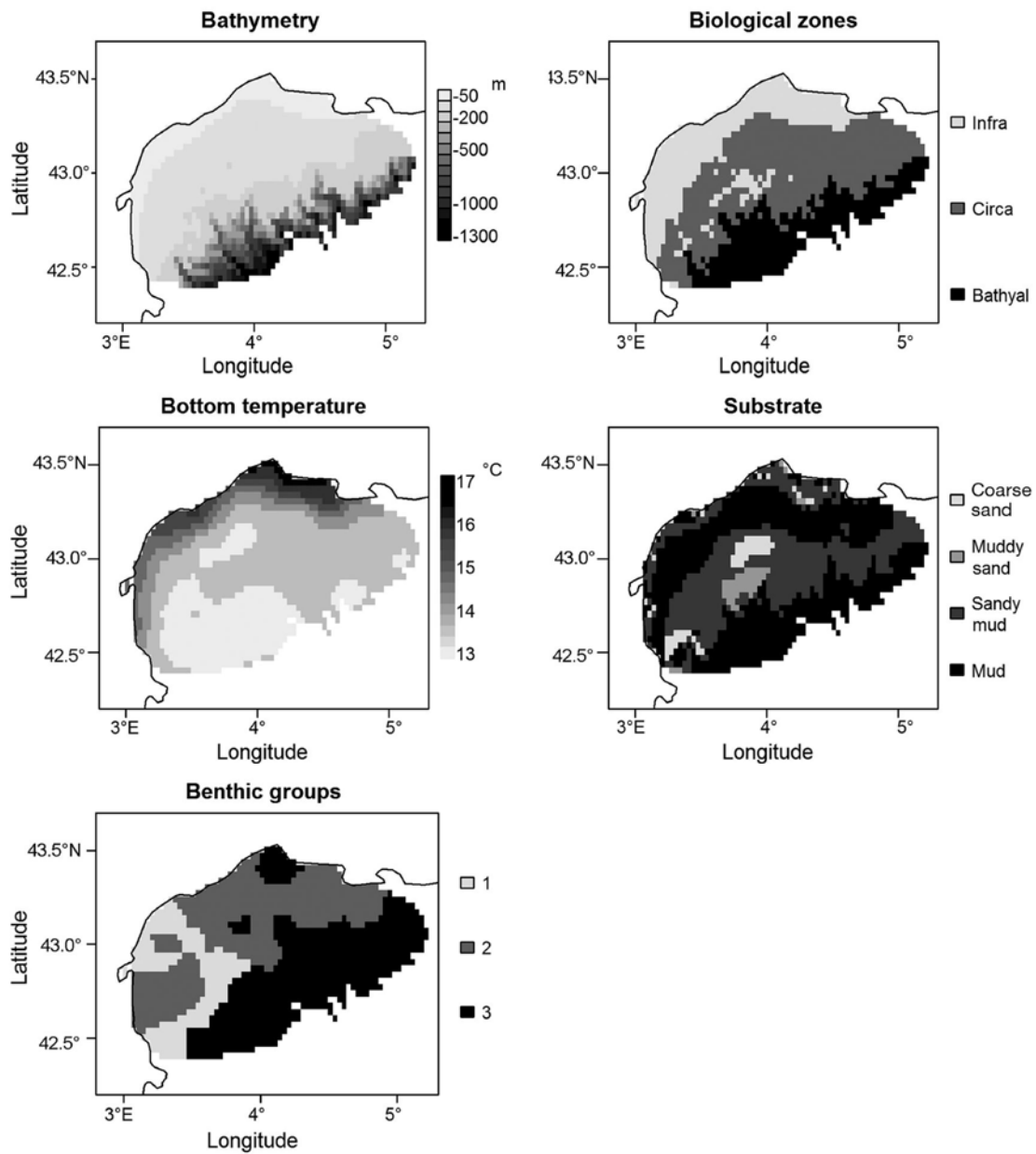
797



798

799 **Figure 1. Study area and sampling sites.**

800 Map of the Gulf of Lions and the 65 sampling sites (identified by crosses) during the
801 whole MEDITS survey (1994-2010). The positions of the sites were set by a stratified
802 sampling scheme, according to bathymetry (contour lines in grey and in meters).

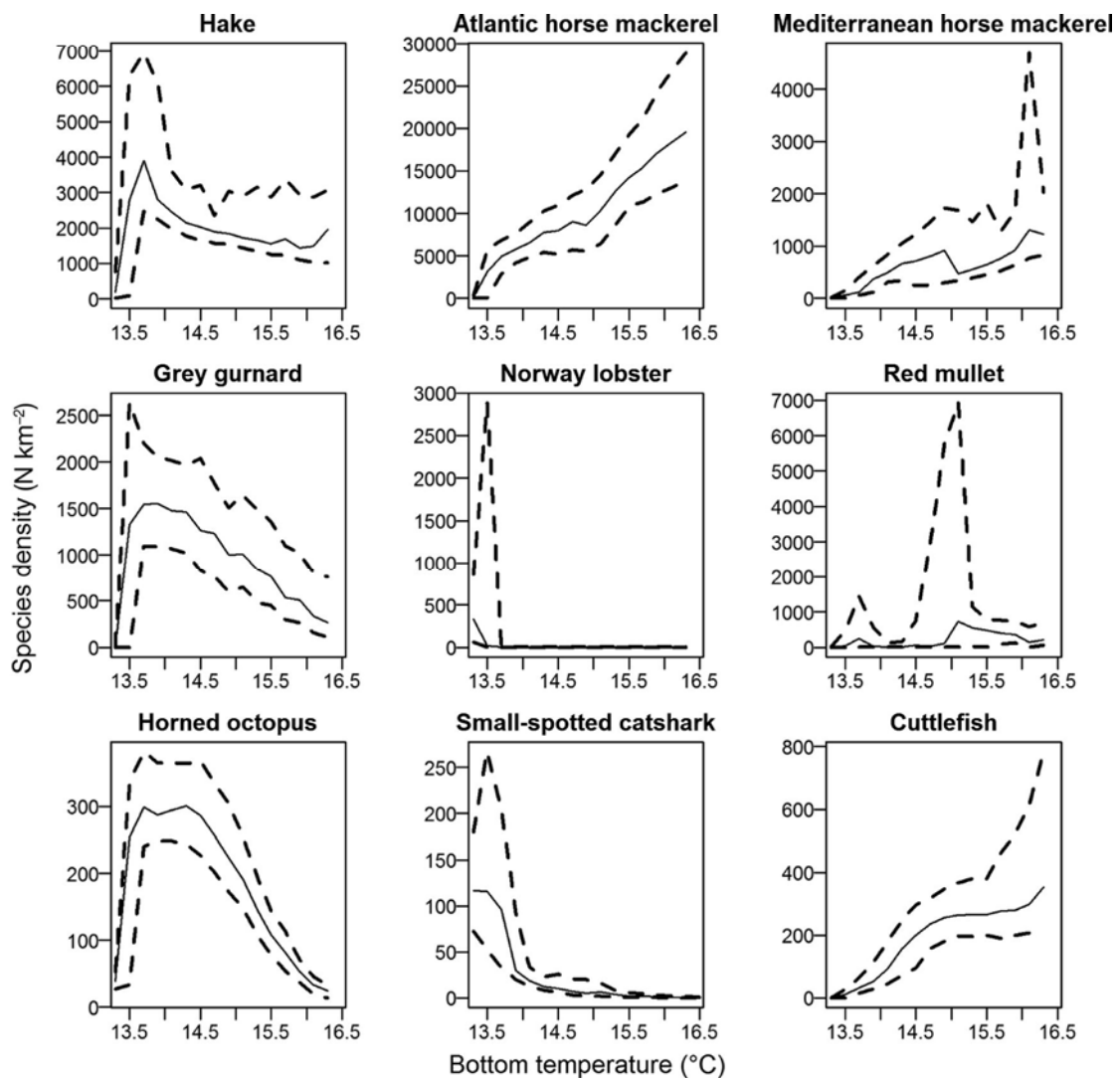


803

804 **Figure 2. External variables**

805 Spatial distributions of the environmental factors considered to model species
 806 distributions. The Biological Zones variable was chosen instead of bathymetry in the
 807 habitat models. Variables are displayed at a 2'x 2' scale.

808

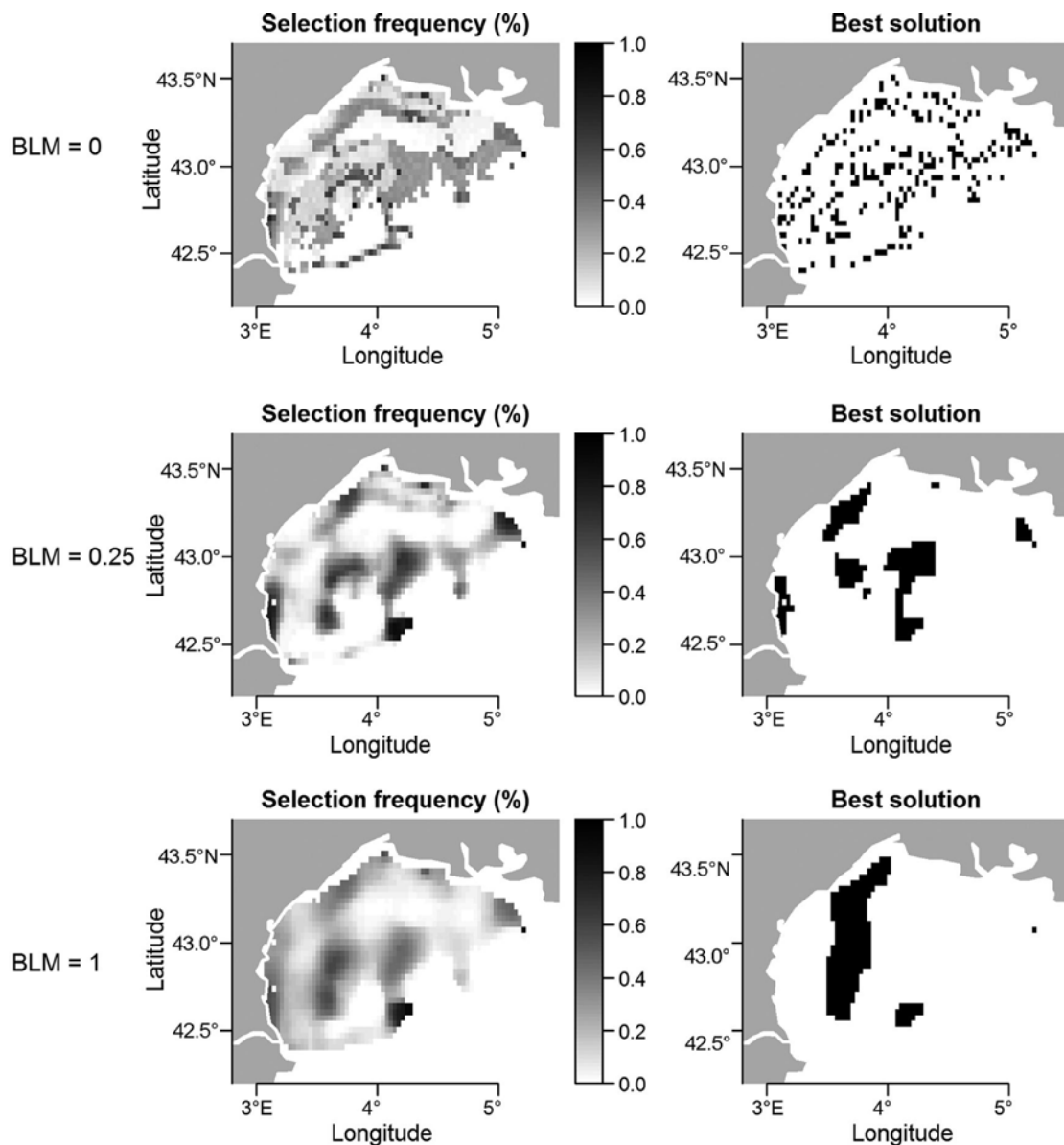


809

810 **Figure 3. Model interpretation: marginal effects.**

811 Marginal response in species density depending on bottom temperature, as predicted by
 812 the model selected by Loo procedure. Distributions of the marginal response were
 813 estimated by bootstrap, solid lines display the means and dashed lines the 10% and 90%
 814 quantiles.

815



816

817 **Figure 4. MPA design**

818 MARXAN outputs, i.e. the minimum set of areas containing 20% of each species
 819 abundance. Results according to three different compacity constraints are displayed:
 820 Boundary Length Modifier (BLM=0) corresponds to no compacity constraint at all and
 821 BLM=0.25 is the best trade-off between area and boundary length (Figure S4). For each
 822 BLM value, the best solution (right) and the selection frequency of each planning unit
 823 (left) among 500 runs are displayed.

824