

---

## Robust identification of potential habitats of a rare demersal species (blackspot seabream) in the Northeast Atlantic

De Cubber Lola <sup>1,2,\*</sup>, Trenkel Verena <sup>1</sup>, Diez Guzman <sup>3</sup>, Gil-Herrera Juan <sup>4</sup>, Novoa Pabon Ana Maria <sup>5</sup>, Eme David <sup>1,6</sup>, Lorance Pascal <sup>1</sup>

<sup>1</sup> DECOD (Ecosystem Dynamics and Sustainability), IFREMER, INRAe, Institut-Agro - Agrocampus Ouest, Nantes, France

<sup>2</sup> Institut de Recherche pour le Développement (IRD), MARBEC (IRD, Ifremer, CNRS, Univ Montpellier), Av. Jean Monnet - Sète, France

<sup>3</sup> AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugarteaz/g, 48395 Sukarrieta Bizkaia, Spain

<sup>4</sup> Centro Oceanográfico de Cádiz (IEO-CSIC), Muelle de Levante s/n, 11006, Cádiz, Spain

<sup>5</sup> Departamento de Oceanografia e Pescas, Universidade dos Açores, 9901-862 Horta, Portugal

<sup>6</sup> RiverLY Research Unit, National Research Institute for Agriculture Food and Environment (INRAE), Villeurbanne, France

\* Corresponding author : Lola De Cubber, email address : [lola.decubber@gmail.com](mailto:lola.decubber@gmail.com)

---

### Abstract :

Species distribution models (SDM) are commonly used to identify potential habitats. When fitting them to heterogeneous, opportunistically collated presence/absence data, imbalance in the number of presence and absence observations often occurs, which could influence results. To robustly identify potential habitats for blackspot seabream (*Pagellus bogaraveo*) throughout its distribution area in the Northeast Atlantic and the western Mediterranean Sea, we used an ensemble species distribution modelling (eSDM) approach, modelling gridded presence–absence data with environmental predictors for two types of occurrence data sets. The first data set displayed the observed unbalanced spatially heterogeneous presence/absence ratio and the second a balanced presence/absence ratio. The data covered the full distribution area, including the European Atlantic shelf, the Azorean region and the Western Mediterranean Sea. Across these regions, populations display variable status. The main environmental predictors for potential habitats were bathymetry and annual maximum SST. The fitted ensemble compromise (eSDM) was projected over the whole grid to create a habitat suitability map. This map exhibited higher probabilities of presence for the balanced-ratio data set. A binary presence–absence map was then generated using optimized presence probability thresholds for four validation indices. Using the true skill statistic to optimize the threshold, the surface areas of the binary presence–absence map was 53% smaller for the balanced data set than for the observed unbalanced data set. However, the choice of validation index had an even greater impact (up to 15 000%). This indicates that studies using opportunistic data for SDM fitting need to pay attention to the effects of presence/absence data imbalance and the choice of validation index to fully evaluate uncertainty.

---

## Highlights

► Ensemble species distribution modelling to identify potential habitat of rare fish. ► Comparison of two occurrence data sets: imbalanced and balanced presence–absence. ► Main predictors: bathymetry and sea surface temperature. ► Higher presence probabilities for the balanced data set. ► Effect of validation index on habitat map larger than data imbalance effect.

## Keywords

*Pagellus bogaraveo*; Species distribution models; Ensemble modelling; Heterogeneous data set; Presence–absence imbalance

## 41 1. Introduction

42 Actual and potential areas of species distribution can be investigated via eco-  
43 logical niche modeling (Soberon and Nakamura, 2009). A species' niche is defined  
44 as a subset of environmental conditions under which populations of a species have  
45 positive growth rates (Soberon and Nakamura, 2009). The habitat is then the ge-  
46 ographical translation of these environmental conditions. The fundamental niche is  
47 the theoretical combination of environmental variable that allows for physiological  
48 processes (feeding, growth, reproduction) to take place (Hutchinson, 1978). Essen-  
49 tial fish habitats, defined as areas or volumes of water and bottom substrates that  
50 provide the most favourable habitats for fish populations to spawn, feed and mature  
51 throughout their full life cycle, are thus the geographical translation of the optimal  
52 part of the fundamental niche of a species (Helaouet and Beaugrand, 2009; Vala-  
53 nis et al., 2008). The realised niche is the subset of environmental conditions the  
54 species is actually using (Soberon and Nakamura, 2009). Species may occur outside  
55 the fundamental niche during migrations. In contrast, the realised niche might be  
56 reduced when densities are low because of intensive predation or fishing (Helaouet  
57 and Beaugrand, 2009). The realised habitat of a species can then be defined as  
58 the geographical translation of the realised niche of a species. It differs from the  
59 species' distribution since all locations displaying the environmental conditions of  
60 the realised niche might not be occupied simultaneously, especially if the species'  
61 distribution is wide.

62 Species distribution models (SDMs) have been used in conservation biology to  
63 describe the habitat distribution of organisms in both marine and terrestrial sys-  
64 tems (Laman et al., 2018; Elith & Leathwick, 2009; Valanis et al., 2008). They are  
65 grounded in the concept of ecological niche (Hutchinson, 1957). They have been  
66 widely used since 2005 and have reached high statistical sophistication in recent  
67 years (Schickele et al., 2020; Jiménez & Soberón, 2020; Robinson et al., 2017). Eco-  
68 logical assumptions implied when using SDMs are that there is niche conservatism  
69 (Crisp et al., 1981) and unlimited dispersal abilities (Wiens et al., 2009) and that  
70 biotic interactions do not influence large-scale distributions (Gleason, 1926; Guisan  
71 and Thuiller, 2005; Schickele et al, 2020). Among the numerous statistical SDMs

72 approaches developed to map fish habitats, ensemble species distribution modelling  
73 (eSDM), also referred to as ensemble niche modelling (Thuillier et al, 2016), which  
74 combines the use of several SDM categories, appears to be a good compromise in  
75 terms of programming skills required, computation time and consistency of the re-  
76 sults (Schickele et al., 2020; Mateo et al., 2019).

77 Data availability is often opportunistic, so that neither the fundamental habitat  
78 nor the realised species' habitat is entirely represented by SDMs. Indeed, the the-  
79 oretical entire range of fundamental environmental conditions of a species is never  
80 fully known and available presence records will never cover the full habitat. Ecol-  
81 ogists thus generally refer to SDM output as potential niche and habitat of the  
82 species of interest (Schickele et al., 2020; Helaouet and Beaugrand, 2009). Several  
83 data filtration and selection processes, as well as physiological prospects (for exam-  
84 ple, optimal environmental ranges for spawning or egg development) can then help  
85 approaching the species' realised or essential habitat (Schickele et al., 2020; Helaouet  
86 and Beaugrand, 2009).

87 Implementing SDMs, especially in the case of widely distributed species such as  
88 the blackspot seabream *Pagellus bogaraveo* (Brünnich, 1768), often requires combin-  
89 ing heterogeneous multiple data sets (Schickele et al., 2020; Fithian et al., 2015).  
90 In the case of presence/absence data, two types of biases have then to be taken  
91 into account. First, detectability might vary among sampling techniques used to  
92 collect data (Kellner and Swihart, 2014). Second, variations in prevalence (i.e. the  
93 number of presence records among sampled points) might reflect primarily varia-  
94 tions in abundance rather than habitat suitability. When data are missing on the  
95 detection probability of sampling techniques, taking into account detection might  
96 not always improve SDM performance (Welsh et al., 2013), and these two biasing  
97 effects (detectability and variations in prevalence) might be difficult to disentangle.  
98 In the case of presence-only data, a common practice is to generate pseudo-absence  
99 data (Schickele et al., 2020). In this case, the number of generated pseudo-absences  
100 is generally set equal to the number of presences (Montgomery, 2005). For actual  
101 presence-absence data, prevalence will vary in space, in particular for large study  
102 areas. This raises questions given spatial predictions from SDMs are known to be

103 sensitive to sample prevalence (Jimenez-Valverde et al., 2021).

104 The general aim of this study was to investigate the potential habitat of the  
105 blackspot seabream and its occupancy level in three regions in the Northeast At-  
106 lantic: Atlantic European shelf, the Azorean region and the Mediterranean Sea.  
107 Occupancy levels were presumed to differ between regions because of the contrasted  
108 population status and variable degree of fishery exploitation. To evaluate the im-  
109 pact of heterogeneous prevalence in the data and obtain robust results we compared  
110 eSDM models using 1) all available presence/absence records, i.e. prevalence varying  
111 over the distribution area of blackspot seabream; 2) the same number of presence  
112 and absence record, i.e. constant prevalence over the distribution area.

## 113 **2. Material and methods**

### 114 *2.1. Case study*

115 The blackspot seabream used to be a widely distributed and abundant species  
116 of the North Eastern Atlantic shelf from the Faroe Islands down to Gibraltar, the  
117 Azores and the Western Mediterranean Sea (Desbrosses, 1932; Sanz-Fernandez et  
118 al., 2019; Pinho et al., 2014; Erzini et al., 2006; Spedicato et al., 2002; D'Onghia et  
119 al., 2010;2012). In fact, it was also referred to as "la dorade commune" (understand  
120 "the common seabream") by French authors in the early 1900s (Desbrosses, 1932;  
121 Olivier, 1928). Increase in fishing effort in the Bay of Biscay (North Eastern Atlantic  
122 shelf) in the 1960s linked to stock declines of other species of fisheries interest such  
123 as hake, associated to its susceptibility to overexploitation, led to a brutal collapse  
124 of this blackspot seabream stock 20 years later in 1975-1985 and low stock size ever  
125 since (see Fig.1a, Lorance, 2011; Guichet et al., 1971; Dardignac, 1988).

126 Blackspot seabream displays three characteristics that make it susceptible to  
127 over-exploitation (Francis and Clark, 2005). First, its biological productivity is low,  
128 individuals reaching 70 cm long in 25 to 30 years and females being mostly the  
129 older individuals since the species is hermaphroditic protandrous, with changing sex  
130 from male to female (Guéguen, 1969, Lorance, 2011). Second, blackspot seabream is  
131 easy to capture during its seasonal migrations because of its aggregative behaviour  
132 (Afonso et al., 2012; 2014).

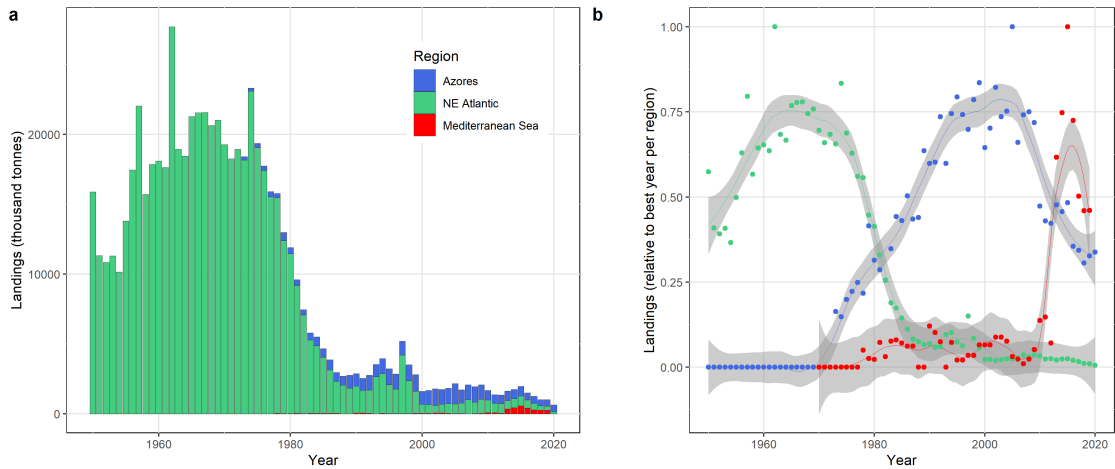


Figure 1: Commercial landings of blackspot seabream between 1950 and 2020 in the three regions investigated in this study as total catch (a) and relative to the highest year per region (b), where dots are individual values and lines are smoothed time trends with 95% confidence intervals. Data before 2000 from Lorance (2011) and from ICES and FAO catch statistics thereafter.

133 Indeed, adults carry out geographic and depth seasonal migrations from coastal  
 134 waters, where they reproduce and where juveniles are found, down to 700 m and  
 135 up to several hundreds of km away from the coast (Morato et al., 2001; Mytilineou  
 136 et al., 2005). Accordingly, in the Bay of Biscay, individuals present to the West of  
 137 Brittany (48°N) were found to overwinter in the Cantabrian Sea (43°N) (Guéguen,  
 138 1974). In Azorean waters, juveniles which are only found in coastal areas migrate to  
 139 isolated seamounts when reaching adult stage, sometimes more than 400 km away  
 140 (Hareide and Garnes, 2001). Lastly, the species has a high commercial value owing  
 141 to its organoleptic quality, comparable to gilthead seabream and sea bass (Rincon  
 142 et al. 2016).

143 Overall, available stock assessments and landings of commercial fisheries sug-  
 144 gest that populations from the European shelf are at low level with small recent  
 145 catches (2018-2021) compared to past levels (anterior to 1980), while populations  
 146 from the Azorean region are in better condition with current fisheries being sustain-  
 147 able (ICES, 2021, Fig. 1). Historically (before 1980), catches from the Northeast  
 148 Atlantic shelf constituted the bulk of landings, reaching up to more than 20000  
 149 t per year (Fig. 1a), with the Bay of Biscay being the main fishing area. In  
 150 comparison, levels of gilthead seabream catches were similar back then, while cur-

151 rently, hake is the most fished species in the Bay of Biscay with around 30000 t  
152 landed per year, followed by monkfish (8000 t per year), sole (3000 t per year) and  
153 seabass (2000 t per year) (Official Nominal Catches 2006-2019. Version 15-10-2021.  
154 Accessed 05-05-2022 via [https://ices.dk/data/dataset-collections/Pages/Fish-catch-](https://ices.dk/data/dataset-collections/Pages/Fish-catch-and-stock-assessment.aspx)  
155 [and-stock- assessment.aspx.](https://ices.dk/data/dataset-collections/Pages/Fish-catch-and-stock-assessment.aspx), ICES). From the 1990s, catches from the Northern At-  
156 lantic came mostly from the Iberian coast and the Strait of Gibraltar and were at  
157 similar level than catches from the Azorean area (Fig. 1a). Reported catches from  
158 the Mediterranean Sea are probably not realistic, because in this region 5000 to  
159 10000 tonnes of fish have been landed as unidentified sparid fish or similar labelling  
160 and this might have comprised catch statistics of blackspot seabream (FAO-GFCM,  
161 2021). Therefore, the increased reported landings in recent years (Fig. 1b) may  
162 be due to improved reporting of landings by species. Quotas as well as other man-  
163 agement measures such as minimum landing size and closed fishing seasons are  
164 implemented in all areas (Pinho et al., 2014; Lorance, 2011). Indeed, fishing has  
165 been shown to be the main factor accounting for variations in the species' stock  
166 abundance, with values reaching up to 73% of the variations in stock abundance  
167 nowadays around Gibraltar (Sanz-Fernandez et al., 2019). In the case of poor stock  
168 status, the species distribution might contract within its essential habitats, that thus  
169 needs to be identified to enable targeted conservation management measures to be  
170 implemented.

## 171 *2.2. Data*

### 172 *2.2.1. Species observations*

173 Presence/absence records of blackspot seabream were compiled from trawling  
174 and longline scientific surveys (EVHOE, SP-NORTH, SP-ARSA, PT-IBTS, MED-  
175 ITS, ARQDAÇO) available on the DATRAS portal ([https://datras.ices.dk/Data\\_](https://datras.ices.dk/Data_products/Download/Download_Data_public.aspx)  
176 [products/Download/Download\\_Data\\_public.aspx](https://datras.ices.dk/Data_products/Download/Download_Data_public.aspx)) or held by national research In-  
177 stitutes, from commercial fisheries data from the Voracera fleet in Gibraltar and from  
178 on-board observations of fishing activities in the Bay of Biscay and the Mediter-  
179 ranean Sea, as well as from the Global Biodiversity Information Facility (GBIF,  
180 <https://www.gbif.org/>) (see Fig.2).

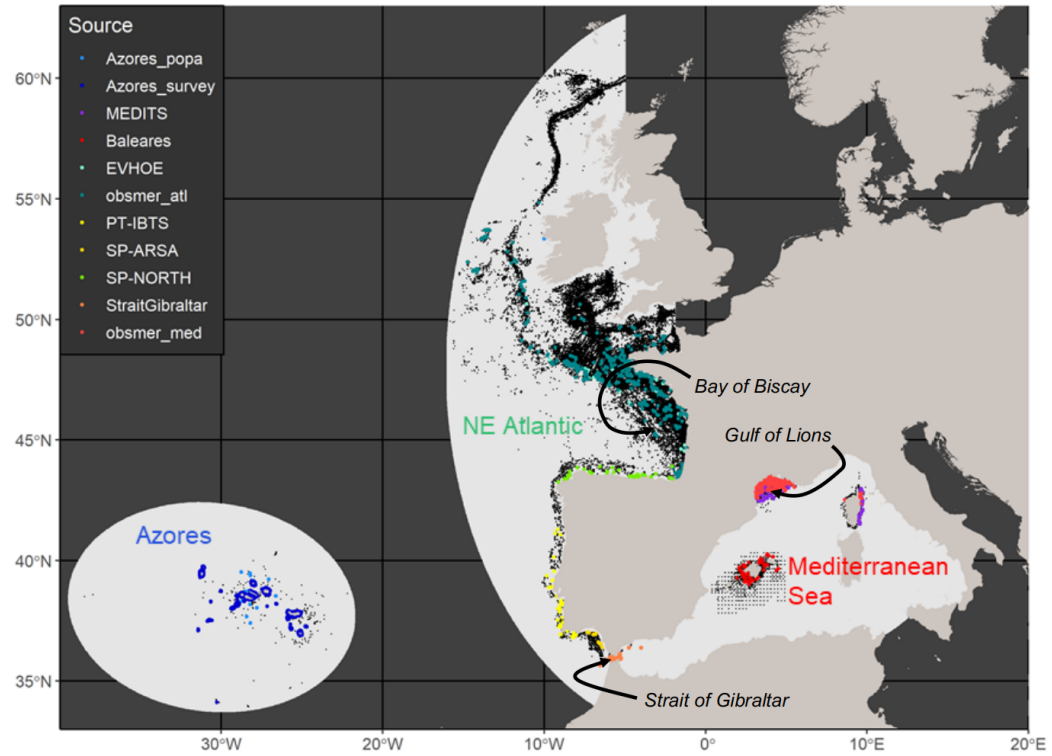


Figure 2: Presence-absence data for blackspot seabream compiled in this study. Black dots represent absence records. Coloured dots correspond to presence records from different data sets: Azores\_popa (Fisheries Observer Program, 1998-2013), Azores\_survey (ARQ-DAÇO longline survey, 1996-2013), Baleares (Marine Biodiversity Atlas of the Balearic Sea, GBIF, 2001-2008), EVHOE (scientific bottom-trawl survey, 1997-2019), MEDITS (scientific bottom-trawl survey, 2004-2019), obsmer\_atl and obsmer\_med (French onboard observation program), PT-IBTS (scientific bottom-trawl survey, 2002-2017), SP-ARSA (scientific bottom-trawl survey, 1996-2019), SP-NORTH (scientific bottom-trawl survey, 2001-2019), StraitGibraltar (Commercial fisheries data, 2009-2011). light grey area is the model domain composed of three regions: Azores, NE Atlantic region and Mediterranean Sea.

### 181 2.2.2. Environmental data

182 Environmental variables consisted of topographic data, sea bottom type and sea-  
 183 water parameters. We extracted bathymetry at a 0.0003° resolution from GEBCO  
 184 ([https://www.gebco.net/data\\_and\\_products/gridded\\_bathymetry\\_data](https://www.gebco.net/data_and_products/gridded_bathymetry_data)) and the  
 185 R terrain function (raster package, Hijmans et al., 2011) enabled the calculation of  
 186 bottom slope. Seabed habitat data were extracted from EMODnet ([https://www.](https://www.emodnet.eu/en/seabed-habitats)  
 187 [emodnet.eu/en/seabed-habitats](https://www.emodnet.eu/en/seabed-habitats)) at a 250 m resolution. Homogenization of substrate  
 188 type according to EMODnet categories among all regions led to 14 sea bottom type  
 189 categories: unknown, rock or other hard substrata, coarse substrate, coarse and



190 mixed sediment, mixed sediment, sediment, sand, sandy mud, muddy sand, sandy  
 191 mud or muddy sand, fine mud or sandy mud or muddy sand, fine mud, *Posidonia*  
 192 *oceanica*, and dead mattes of *Posidonia oceanica*. Monthly values of Sea Surface  
 193 Temperature (SST), bottom temperature, surface current velocity and salinity be-  
 194 tween January 1994 and December 2018 at a 0.083° resolution were extracted from  
 195 Copernicus Marine Service (GLOBAL\_REANALYSIS\_PHY\_001\_030 product,  
 196 <https://resources.marine.copernicus.eu/>). An overview of environmental variables  
 197 investigated is provided in Sup. Mat. 1. The mean, maximum, minimum and stan-  
 198 dard deviation of environmental variables were computed for each grid cell (n =  
 199 6465).

### 200 2.3. Methods

201 The general workflow used in this study is presented in Fig. 3. Model reporting  
 202 was done following recommendations by Zurell et al. (2020).

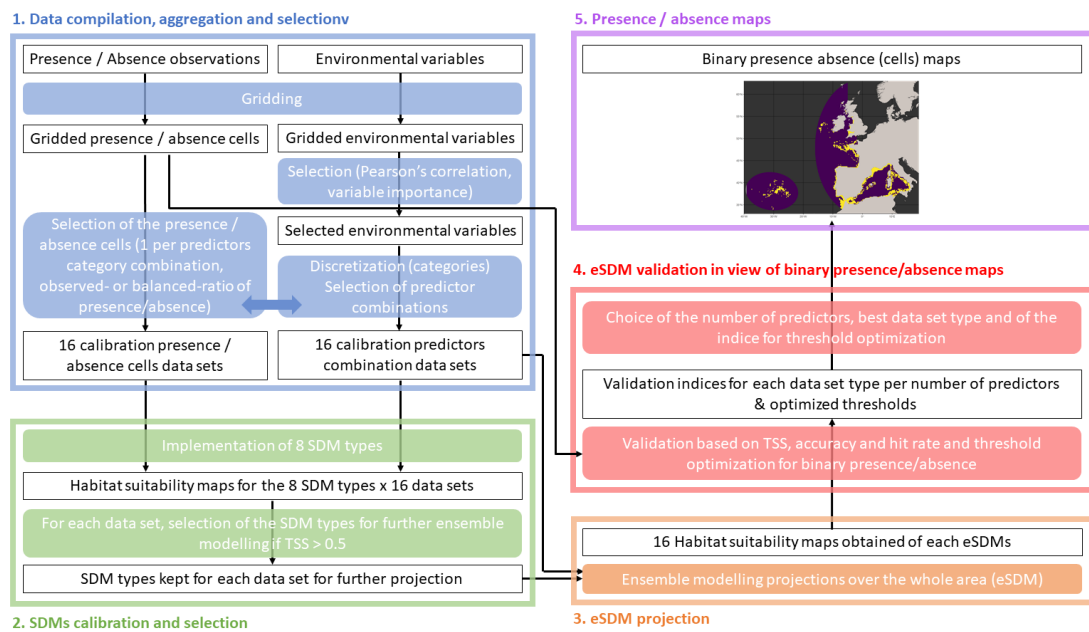


Figure 3: General workflow of the procedure used for identifying blackspot seabream habitats: (1) compilation, gridding and selection of environmental and occurrence data, (2) statistical Species Distribution Models (SDM) calibration (for each category) and selection, (3) Projection of the ensemble (eSDM) model (habitat suitability map), (4) Validation of projections based on maximization of various indices and determination of threshold value for (5) binary presence/absence mapping.

203 *2.3.1. Study area*

204 Three regions were modelled corresponding to the three main areas of distribution  
205 of the blackspot seabream: The Northeast Atlantic shelf, the Azorean region and the  
206 western Mediterranean basin (Fig. 2). A spatial grid of  $0.1^\circ \times 0.1^\circ$  cells (44851 in  
207 total) was created covering the three regions. This resolution was chosen accounting  
208 for the need for local and general overviews of species habitat, as well as computation  
209 time.

210 *2.3.2. Occurrence data compilation, gridding and selection*

211 For each grid cell with data ( $n = 6465$ ), the number of presence and absence  
212 records was counted. The occurrence data set displayed over- and under-sampled  
213 areas (Fig. 2). In order to take into account this heterogeneity, occurrence records  
214 were compiled into presence/absence records: cells with at least one presence were  
215 considered as presence cells, cells with only absence records were considered as ab-  
216 sence cells and the remaining cells were considered as non-sampled cells and were not  
217 used for model fitting. In order to maintain an homogeneous distributions of sam-  
218 pled cells along environmental gradients, continuous environmental variables were  
219 discretised into 60 classes each (which appeared to be a good compromise to create  
220 relevant classes for all environmental variables) over the whole grid domain. When  
221 a given combination of environmental variable classes corresponded to several pres-  
222 ence cells, only one presence cell was kept, similarly for absence cells. When a given  
223 combination corresponded to several presence and absence cells, one presence and  
224 one absence cell was kept. Thus the number of data points used for model fitting  
225 depended on the environmental predictors included in a particular model (Table 2).

226 Since the ratio between the number of presence and absence grid cells still varied  
227 among regions after the gridding process, two approaches were tested. In the first  
228 case (observed-ratio data set), all available presence / absence grid cells were kept  
229 in the next steps of the analysis. In the second case (balanced-ratio data set), the  
230 same number of absence and presence grid cells was used for all regions by randomly  
231 selecting absence grid cells among all available absence cells. The second approach  
232 corresponds to the common practice for presence-only data for which pseudo-absence

233 data are created (Montgomery, 2005)(See Fig. 3, step 1).

### 234 *2.3.3. Selection of environmental predictors*

235 Given the observation of heterogeneous responses to some of the environmental  
236 variables according to the region (Azores, Atlantic shelf and Mediterranean Sea, see  
237 Sup. Mat. 2), a categorical predictor was added for region. To identify the most par-  
238 simonious environmental data set explaining the blackspot seabream distribution,  
239 pairwise correlations between all environmental variables were investigated with a  
240 Pearson's correlation test using as correlation threshold  $r > 0.7$  (Schickele et al.,  
241 2020; Dormann et al. 2013). When several environmental variables were highly cor-  
242 related, we retained the environmental variable with the highest relative importance  
243 (Schickele et al., 2020; Leroy et al., 2014)(Sup. Mat. 3). As the previous selection  
244 step led to a high number (9) of remaining environmental predictors for both occur-  
245 rence data sets, multiple models were fitted with decreasing number of predictors  
246 (9 to 2 predictors), removing sequentially the predictor with the smallest relative  
247 importance to balance model fit and model complexity (Meynard et al., 2019) (See  
248 Fig. 3, step 1). Next, for each of the eight combinations of predictors and each  
249 occurrence data set type (observed-ratio and balanced-ratio), one occurrence data  
250 set was created for calibration (16 data sets).

### 251 *2.3.4. Species distribution model categories and settings*

252 The following eight SDM categories were implemented in R using the BIOMOD2  
253 package (Thuillier et al., 2003;2016) with default parameter settings: generalised lin-  
254 ear model (GLM), generalised boosting model (GBM), generalized additive model  
255 (GAM), artificial neural network (ANN) model, flexible discriminant analysis (FDA),  
256 random forest (RF), classification tree analysis (CTA) and surface range envelope  
257 (SRE) model (Valanis et al., 2008; Thuiller et al., 2009; Albouy et al., 2012; Clair-  
258 baux et al., 2019; Pecchi et al., 2019). For each SDM category and data set, a 3-fold  
259 cross validation procedure was performed.

260 *2.3.5. SDM category selection and compromise*

261 We used the True Skill Statistics index (TSS, Allouche et al., 2006) to quantify  
262 the performance of each of the eight fitted SDMs for each calibration data set (See  
263 Fig. 3, step 2). It was calculated as  $TSS = \text{sensitivity (proportion of presence cells}$   
264  $\text{classified correctly}) + \text{specificity (proportion of correct absence cells)} - 1$ . A SDM was  
265 selected for ensemble modelling if it had  $TSS > 0.5$ . A compromise (the ensemble  
266 Species Distribution Model, eSDM) of the presence probability was then calculated  
267 as the mean of probabilities of retained SDMs weighted by their TSS value. Presence  
268 probability uncertainty was quantified using the coefficient of variation from cross-  
269 validation results. Individual SDM response curves to environmental predictors as  
270 well as the eSDM resulting presence probabilities according to the different predictors  
271 (Schickele et al., 2020) are presented in Sup. Mat. 4 and 5.

272 *2.3.6. Binary habitat maps and predictors selection*

273 To define habitat suitability maps, for each of the 16 calibrated ensemble models  
274 (2 data sets x 8 predictors combination with 2 to 9 predictors), presence probabilities  
275 were projected over the whole domain, including the cells not included in the cali-  
276 bration process (extrapolation for the non sampled cells). Then, in order to create  
277 binary presence/absence maps, the habitat suitability maps (presence probabilities)  
278 were compared to all observed presence and absence compiled cells (not only those  
279 used for model fitting). Threshold values for binary projections of presence and  
280 absence were calculated based on a set of indices as recommended by Robinson et  
281 al. (2017), namely the hit rate (proportion of correctly classified presence cells),  
282 the True Skill Statistics (TSS, Allouche et al., 2006), the CBI (computed with the  
283 `ecospat.boyce` function of the `ecospat` package on R, Hirzel et al. 2006), and the  
284 overall accuracy (sum of the proportion of correctly classified presence and absence  
285 cells, Allouche et al., 2006). More precisely, each validation index was maximized  
286 varying threshold values for binary projections of presence and absence (from 0 to 1  
287 with a 0.001 interval) with the `optimize` function in R. Each maximized index value  
288 thus corresponded to a distinct optimized threshold value.

289 For each data set type, the best calibrated eSDM was chosen as the one showing

290 the lowest number of predictors and the highest validation index values. To combine  
 291 validation index values, their values were re-scaled between 0 (lowest index value  
 292 obtained across eSDM outputs with 2-9 predictors) and 1 (highest index value). In  
 293 the case of the balanced-ratio data set, as absence cells selection might have impacted  
 294 ensemble model performance in the previous steps, 10 data sets with the chosen  
 295 number of predictors and varying absence cells were randomly selected and the  
 296 average of habitat suitability maps (re-calibration) was used for further validation.  
 297 Lastly, for each data set type and validation index, the surface of potential habitat  
 298 was calculated.

### 299 3. Results

#### 300 3.1. Data characteristics

301 In total, 106 457 occurrence records were compiled, among which 6465 presence  
 302 records, corresponding to 782 cells where the species was present and 5683 cells  
 303 where it was recorded as absent (Fig. 2, Table 1).

Table 1: Summary of blackspot seabream occurrence data by region. Domain per region as in Fig. 2. Records correspond to point observations of presence or absence of the species while cells correspond to grid cells of the domain where one or several presence or absence point observation was made. One presence observation was sufficient to qualify as presence cells.

	Total area			Atlantic			Azores			Mediterranean		
	records	cells	rec./cell	records	cells	rec./cell	records	cells	rec./cell	records	cells	rec./cell
N presence	6928	782	8.8	639	389	1.6	4872	165	29.5	1417	228	6.2
Proportion		1.7 %			1.7 %			1.8 %			1.9 %	
N absence	99529	5683	17.5	74556	4626	16.1	13906	376	37	11067	681	16.3
Proportion		12.7 %			20 %			4%			5.6 %	
N sampled	106457	6465	16.5	75196	5015	15	18778	541	34.7	12484	909	13.7
Proportion		14.4 %			21.6 %			5.8 %			7.4%	
N presence / N sampled	6.5 %	<b>12 %</b>		0.8 %	<b>7.8 %</b>		25.9 %	<b>30.5 %</b>		11.4 %	<b>25 %</b>	
N cells per region		44851			23209			9392			12250	

304 The proportion of presence cells among sampled cells was highly heterogeneous  
 305 between regions, as well as the number of records per cell (respectively 7.8 to 30.5 %  
 306 and 14 to 35 records per cell, see Table 1).

#### 307 3.2. eSDM outputs and habitat suitability maps

308 After predictor selection using pairwise correlation analysis, nine predictors were  
 309 retained for the ensemble modelling procedure: region (NE Atlantic, Azores, Mediter-

310 ranean Sea), sea bottom type, mean and standard deviation of bathymetry (m), an-  
311 nual maximum Sea Surface Temperature (SST, °C), annual mean and minimum ab-  
312 solute current velocity (m.s<sup>-1</sup>), annual standard deviation of salinity (‰) and annual  
313 standard deviation of bottom temperature (°C) (Sup. Mat. 3). Mean bathymetry  
314 and maximum SST had the highest explanatory power for both data sets, both vari-  
315 ables contributing equally (Table 2). The main difference between ensemble models  
316 for the two data set types was that for the observed-ratio data set (heterogeneous  
317 ratios of presence and absence over the model domain), the categorical predictor "re-  
318 gion" had a high relative importance, which was not the case for the balanced-ratio  
319 data set (balanced number of presence and absence grid cells in each region) (Table  
320 2, Sup. Mat. 6).

321 SDM categories ANN, GBM and RF were selected in most cases (TSS > 0.5),  
322 while the selection of the other SDM categories varied according to the number of  
323 predictors and the data set type used (Table 2, Sup. Mat. 6). TSS values were higher  
324 for the observed-ratio data set for all models (two to nine predictors), with values  
325 around 0.7 for the balanced-ratio data set, and around 0.9 for the observed-ratio data  
326 set (Table 2, Sup. Mat. 6). As a result of the data selection procedure, the data set  
327 size decreased with decreasing number of predictors, leading to low data set sizes  
328 for models with less than five predictors (Table 2). Comparing projected habitat  
329 suitability maps (presence probabilities per grid cell over the whole domain) between  
330 data set types, it appeared that a balanced ratio between presence and absence  
331 data led to overall higher presence probabilities and hence a much wider potential  
332 habitat in the NE Atlantic region and in the Mediterranean Sea region compared to  
333 the results obtained with the observed-ratio data set, while the projected presence  
334 probabilities appeared rather similar for the Azores region (Figs. 4 and 5). The  
335 visual difference was confirmed by the mean projected presence probability over the  
336 whole domain being  $0.16 \pm 0.19$  for the balanced-ratio data set and only  $0.07 \pm 0.09$   
337 for the observed-ratio data set (Fig. 4).

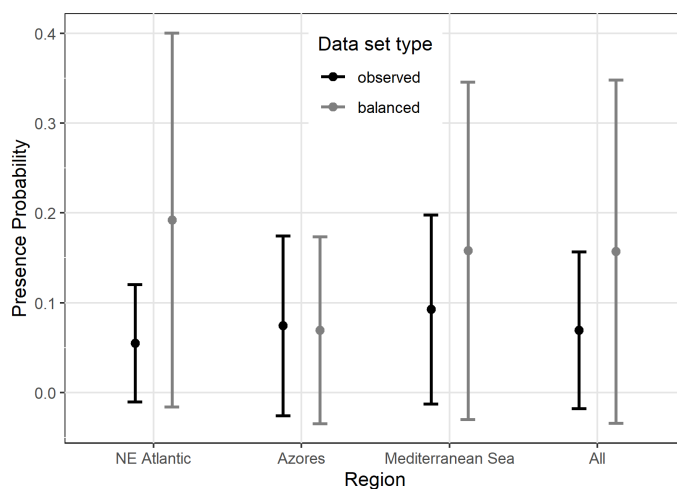


Figure 4: Presence probabilities (mean and sd) of the blackspot seabream per region and on the whole grid according to the data set type used for ensemble species distribution modelling with 6 predictors.

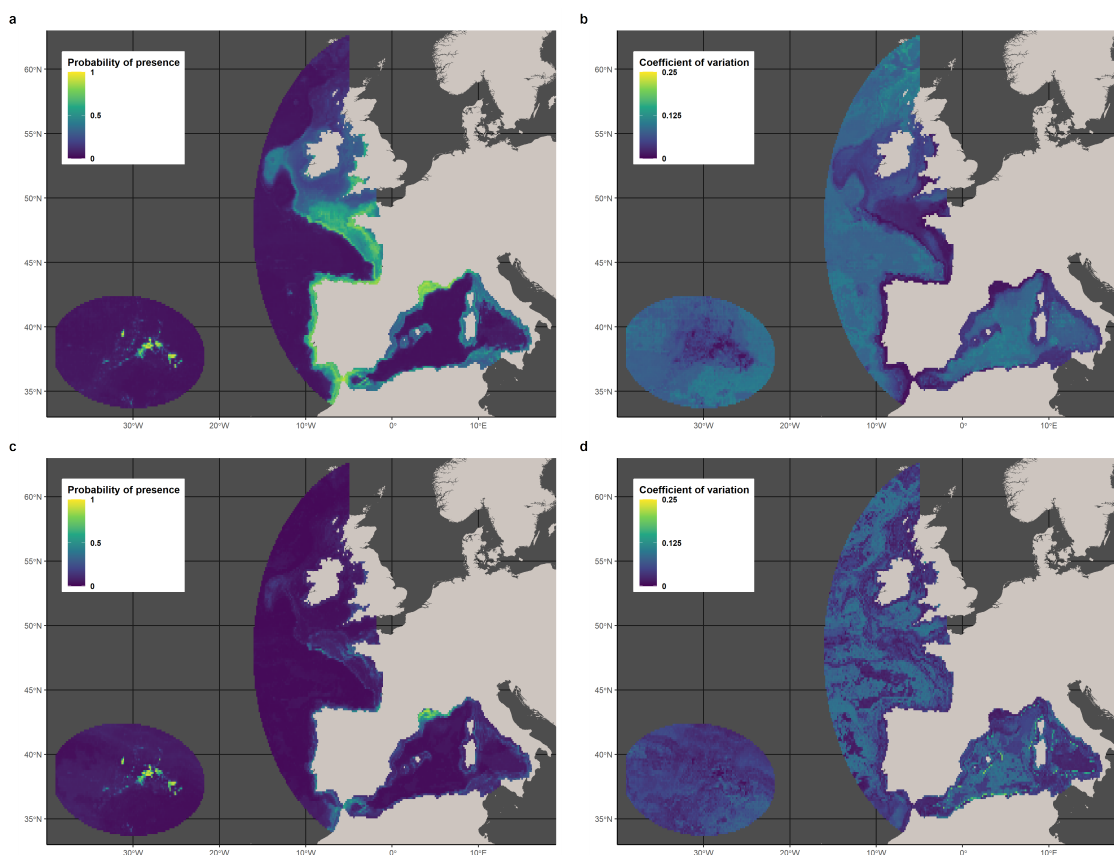


Figure 5: Habitat suitability maps for blackspot seabream for the main distribution area (a, c) and associated coefficients of variation (b, d) obtained with an ensemble species distribution model performed using a data set with a balanced number of presence and absence data (balanced-ratio, a, b) and a data set with the observed and heterogeneous number of presences and absences (observed-ratio, c, d), for models including six predictors (see Table 2 and Sup. Mat. 6).

Table 2: Summary of the 16 calibrated ensemble species distribution models implemented in this study with 8 different numbers of predictors for the observed-ratio data set type: predictors used and their relative importance (proportion of explained variance), characteristics of the data set used, SDMs included in eSDM and overall performance.

	2 predictors	3 predictors	4 predictors	5 predictor	6 predictors	7 predictors	8 predictors	9 predictors
<b>Environmental variables' importance</b>								
Mean bathymetry	50%	49 %	50%	37%	30%	31%	26%	27 %
Max SST	50%	34%	18%	25%	31%	25%	27%	26%
Sd bathymetry	-	17%	18%	20%	18%	16%	12%	11%
Sd bottom temperature	-	-	-	8%	8%	9%	7 %	8%
Sd salinity	-	-	-	-	5%	6 %	5 %	5%
Min current velocity	-	-	-	-	-	3%	4%	2%
Mean current velocity	-	-	-	-	-	-	-	2 %
Sea bottom type	-	-	-	-	-	-	1%	1%
Region	-	-	11%	13 %	9%	10%	18%	17%
<b>Occurrence data set</b>								
Data set size	474	1510	1696	2620	3280	3357	3752	4163
Number of presence cells	8	161	161	372	470	475	536	576
Number of absence cells	466	1349	1535	2248	2810	2882	3216	3587
<b>SDM types and eSDM performance</b>								
Selected models (TSS > 0.5)	ANN, FDA, GAM, GBM, GLM, RF	ANN, FDA, GAM, GBM, GLM, RF	ANN, CTA, FDA, GBM, GLM, RF, SRE	ANN, GBM, GLM, RF	ANN, GBM, RF	ANN, GBM, RF	ANN, GBM, RF	ANN, GBM, RF
True Skill Statistic (mean $\pm$ sd across SDMs)	0.996 $\pm$ 0.003	0.791 $\pm$ 0.013	0.826 $\pm$ 0.013	0.867 $\pm$ 0.033	0.916 $\pm$ 0.030	0.895 $\pm$ 0.039	0.904 $\pm$ 0.043	0.894 $\pm$ 0.053



338 Differences between projected presence probabilities obtained with the balanced-  
 339 ratio data set and the observed-ratio data set were the highest in the NE Atlantic  
 340 (with respective values of  $0.19 \pm 0.21$  and  $0.06 \pm 0.07$ ) and the lowest in the Azores  
 341 (with both values equal to  $0.07 \pm 0.10$ ) (Fig. 4). For both data set types, coefficients  
 342 of variation of presence probabilities did not exceed 0.25 (Fig. 5).

343 Overall, the choice of data set type impacted presence probability values, but  
 344 did not impact consistently which predictors were selected, except for the region  
 345 predictor, nor the general shape of the response curve for each predictor, except for  
 346 maximum annual SST for which the relationship was dome-shaped for balanced-  
 347 ratio data set and more flat for the observed-ratio data set (Table 2 and Sup. Mat.  
 348 4).

### 349 3.3. Binary habitat maps

#### 350 3.3.1. Predictors

351 Comparison of the three validation index values between eSDMs using 2 to 9  
 352 predictors normalized between 0 and 1 (0 being the smallest index value across the  
 353 set of predictors and 1 being the highest) led to select the eSDM with six predictors  
 354 for both data set types. Indeed, these models displayed among the best index values  
 355 for the smallest number of predictors (Fig. 6).

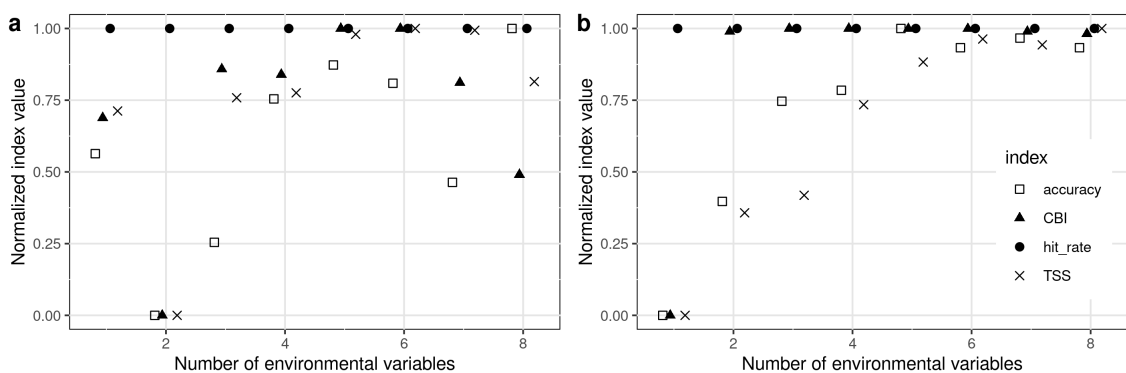


Figure 6: Normalized values of validation indices (accuracy, Allouche et al., 2006; hit rate, Allouche et al., 2006; True Skill Statistics, Allouche et al., 2006, Continuous Boyce Index, Hirzel et al. 2006) of the models according to the number of predictors used, for the balanced-ratio data set type (same number of absences and presences cells) (a) and the observed-ratio data set type (all records) (b).

356 The best predictors for the balanced-ratio data set were mean bathymetry, annual  
 357 maximum SST, standard deviation (sd) of bathymetry, sd of mean annual bottom

358 temperature, sd of mean annual bottom salinity and minimum annual absolute cur-  
 359 rent velocity. For the observed-ratio data set, predictor region was selected instead  
 360 of minimum annual absolute current velocity (Table 2, Sup. Mat. 6). The region  
 361 predictor had a 9% contribution to the explained variance and, compared to the  
 362 balanced-ratio data set, contribution of other variables was lower for bathymetry  
 363 and higher for all other variables.

### 364 3.3.2. Indices and threshold optimization

365 Validation index values (accuracy, hit rate, CBI and TSS) were slightly higher  
 366 for the observed-ratio data set than for the balanced-ratio data set (Sup. Mat. 8).  
 367 Variations in estimated habitat area between validation indices were higher than  
 368 between data set types (Fig. 7). Overall, the choice of threshold value used for  
 369 transforming presence probabilities into binary habitat maps strongly influenced  
 370 results for the whole area and for each of the regions (Fig. 8). For regional habitat  
 371 area estimation, separate threshold values were obtained by maximising index values  
 372 regionally. Maximizing the hit rate led to the lowest threshold values for the whole  
 373 area and each region, and subsequently the largest habitat areas.

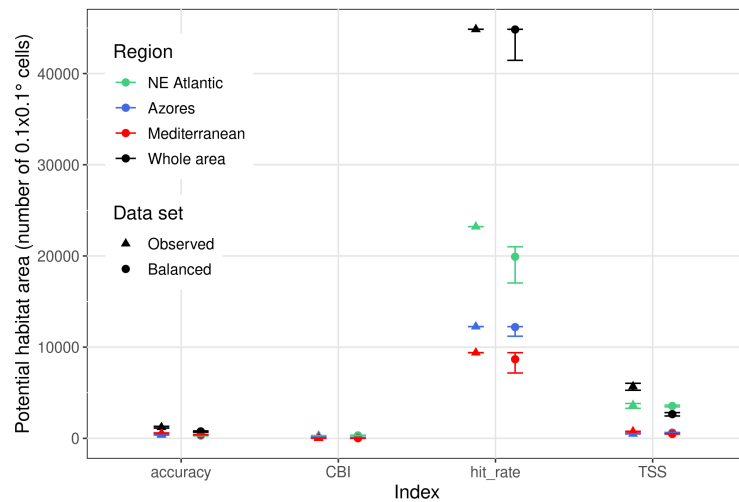


Figure 7: Blackspot seabream potential habitat area using the best model (6 predictors) and different validation indices for setting the presence/absence threshold value: accuracy, hit rate, Continuous Boyce Index and True Skill Statistics. Results for observed- and balanced-ratio data sets, when models were optimized for the whole model domain (black), the NE Atlantic region (green), the Azores region (blue), the Mediterranean Sea region (red).

374 CBI and accuracy led to the smallest habitat areas, especially for the NE Atlantic

375 region for which the observed presence / absence ratio was low. For the whole area,  
 376 averaged across the two data types, the habitat areas for hit rate and TSS were  
 377 14849% and 1280% larger than for CBI respectively. Comparing results between the  
 378 two data sets showed that the balanced data set led to smaller habitat areas for the  
 379 majority of indices for the whole area and each of the three regions (Figs. 7 and 8).  
 380 For the whole area, the difference  $((\text{balanced-observed})/\text{observed})$  ranged between  
 381 -53% for TSS and 4% for CBI. For the Azores the difference ranged between -3% for  
 382 accuracy and 27% for TSS, for the Mediterranean Sea between -36% for TSS and  
 383 0% for CBI, and for the NE Atlantic region between -28% for accuracy and 23 %  
 384 for CBI. Thus, the choice of validation index had a much greater impact than the  
 385 data set type, though the later was also important for certain indices.

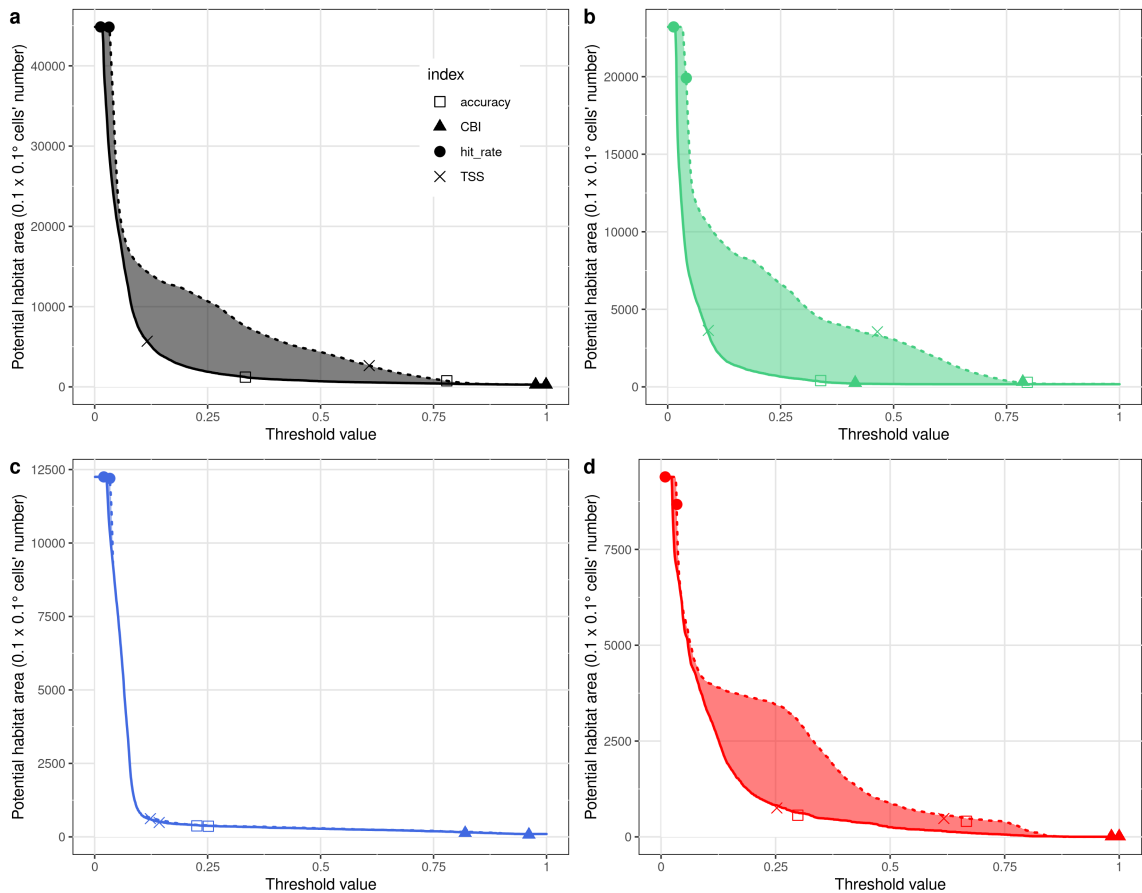


Figure 8: Area of the potential habitat of the blackspot seabream as a function of the threshold value used for binary presence/absence predictions for the balanced-ratio data set (dotted line) and the observed-ratio data set (solid line) and results obtained using different validation indices (accuracy, hit rate, CBI and TSS) for the whole area (a), NE Atlantic region (b), Azores region (c) and Mediterranean Sea region (d).

### 386 3.3.3. *Binary habitat maps*

387 Visual comparison of binary (presence / absence) habitat maps (threshold opti-  
388 mized according to TSS index) between data set types showed that for the balanced-  
389 ratio data set some locations from which the species had been reported were missed,  
390 particularly to the West of Ireland and along the Mediterranean coast (Figs. 2 and  
391 9). Overall, it appears that the potential habitat of the species covers a large area  
392 around seamounts in the Azores region, a wide area on the NE Atlantic shelf and a  
393 narrower area on the Mediterranean shelf (Fig. 9).

## 394 4. Discussion

395 In this study, to obtain robust estimates of the potential habitat of blackspot  
396 seabream across its wide distribution area exhibiting varying exploitation status, an  
397 ensemble species distribution modelling approach and two data sets with different  
398 prevalence levels were used. The identified potential habitats included islands con-  
399 tours and seamounts in the Azores region, the NE Atlantic shelf south of 48°N, with  
400 smaller areas further North, and the Northern shores of the western Mediterranean  
401 Sea, with more extended areas in the Strait of Gibraltar, in the Gulf of Lions and  
402 along the Italian coast.

403 Potential habitats of blackspot seabream were best explained by bathymetry  
404 (down to 700 - 1000 m) and SST (annual maximum SST generally greater than  
405 16°C), as well as bottom temperature, salinity and region as secondary predictors,  
406 independent of the prevalence level in the data set except for region. Further, the  
407 general shape of the response curves for each predictor were similar for the two data  
408 sets except for max SST (see response curves in Sup. Mat. 4). Other studies have  
409 reported the same main factors influencing the species' distribution and abundance,  
410 with occurrences reported in areas with bottom depths between 100 m and 700 m  
411 (Mytilineou et al., 2014; Burgos et al., 2013; Menezes et al., 2013; D'Onghia et  
412 al., 2010; Gueguen, 1974) and environmental conditions linked to temperature and  
413 salinity influencing stock variations (Sanz-Fernandez et al., 2019).

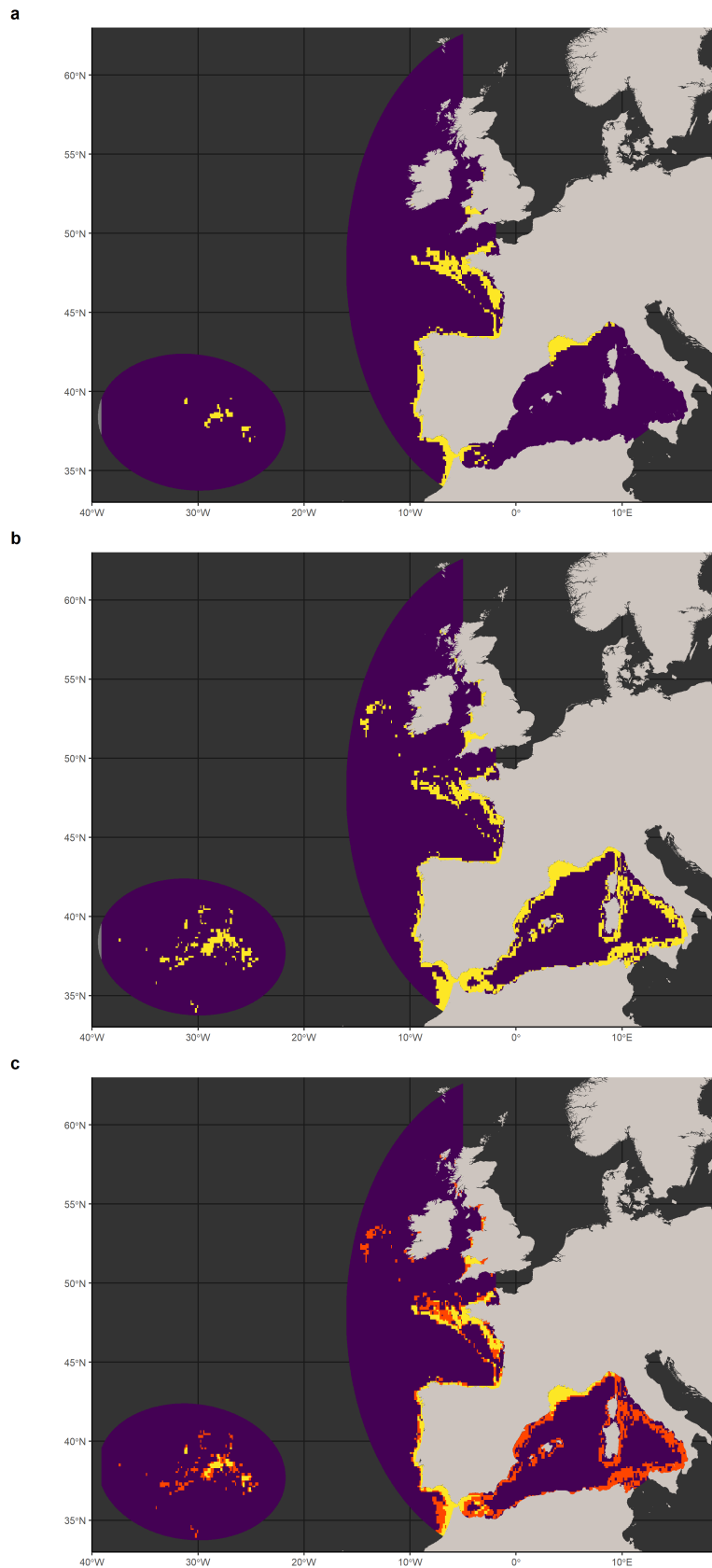


Figure 9: Map of estimated potential habitats with six predictors (yellow) for blackspot seabream over its whole area of distribution using the balanced-ratio data (a) or the observed-ratio data set (b) maximizing the TSS index for presence threshold estimation, and difference between the two potential habitat maps (c, red).

414 The general importance of the region predictor can be explained by regional  
415 differences in biophysical conditions while the difference in importance of this pre-  
416 dictor for the two prevalence data sets is most likely caused by differences in the  
417 prevalence in the observed-ratio data sets (Table 2). These regional discrepancies  
418 in prevalence might be explained by the differences in population status, with the  
419 Northeast Atlantic population being most depleted. It is less likely caused by dif-  
420 ferences in detectability by the employed sampling methods, which could however  
421 be explored by comparing several methods for the same area and using models in-  
422 tegrating detectability (Fithian et al., 2015; Fletcher et al., 2016; 2019).

423 In the Bay of Biscay, which is in the center of the Northeast Atlantic region, the  
424 eSDM suggested that the potential habitat covered a large part of the continental  
425 shelf. Indeed, the species used to occur over most of this shelf before the population  
426 collapsed in the late 1970s. In the Azorean region, which has been termed an  
427 oceanic seamount ecosystem (Silva and Pinho, 2007), the species occurs around  
428 islands, with juveniles distributed in near coast habitats as in the Bay of Biscay  
429 and adults spread from the coast to island slopes and isolated seamounts; juveniles  
430 never occur at sea mounts (Pinho et al., 2014). Depth was identified as the main  
431 factor for explaining blackspot seabream abundance and size composition on Azorean  
432 seamounts (Santos et al., 2021). In the same area, Morato et al. (2001) showed that  
433 the species feeds on both pelagic and benthic preys and suggested that its food supply  
434 on seamounts may depend upon oceanic production that drifts past seamounts,  
435 which in turn makes bottom slope a factor for the species distribution as current  
436 strength and therefore the amount of prey advected increase with slope. At the scale  
437 of the environment perceived by individual fish, similar hydrological conditions,  
438 in particular strong and variable tidal currents resulting from the interaction of  
439 oceanic water masses (Koslow, 1996; Lorange et al. 2002) prevail at seamounts  
440 and along the continental slope, where blackspot seabream also occurs. At this fine  
441 scale, the species shows diel vertical migrations (Afonso et al., 2012), which may  
442 suggest a behavior similar to that of deeper living "seamount aggregators" which  
443 share with blackspot seabream a high lipid content and strong swimming abilities,  
444 which are both related to high metabolism (Koslow, 1996). Overall, the habitat

445 characteristics of fish displaying this type of behavior include sloping sea bottom  
446 and related variations in current speed, which generate also temperature variations  
447 and is in-line with habitats variables found to be predictors of *blackspot seabream*  
448 presence-absence in this study.

449 The ratio between presence and absence cells in the data impacted strongly the  
450 projected probabilities of presence. Imposing a balanced-ratio between the number of  
451 presence and absence cells for model calibration led to higher presence probabilities  
452 on average compared to using the observed-ratio data set with a large majority of  
453 absence cells ( $0.16 \pm 0.19$  and  $0.07 \pm 0.09$  respectively). Thus, including primarily  
454 absence cells in model calibration reduced estimated presence probabilities.

455 The habitat suitability map for the balanced-ratio data set indicated wide po-  
456 tential habitats over the NE Atlantic shelf and along the western Mediterranean  
457 and Azores coasts, while potential habitats were smaller using the observed-ratio  
458 data set. The habitat size was reversed between the two data sets for the corre-  
459 sponding presence/absence maps, with larger binary habitats for the observed-ratio  
460 data set due to a smaller optimized threshold value. These binary maps obtained  
461 by applying a optimized presence probability threshold are probably closer to the  
462 realised habitat of the species as it involved the full presence/absence data to set the  
463 threshold values. The binary presence habitat was notably smaller over the Euro-  
464 pean shelf, where one regional population is depleted. The past high abundance of  
465 the species throughout the Bay of Biscay (Olivier, 1928; Desbrosses, 1932; Guichet  
466 et al., 1971) indicates that this area was suitable for the species 100 to 50 years ago.  
467 The increase of fishing effort on the species at the same period (Lorance, 2011) must  
468 indeed have been the main trigger for the species' decreased abundance, leading to  
469 a concentration of the remaining individuals into the most suitable habitats in this  
470 region (along the coast around isobath -100m and on the continental slope). The  
471 present study suggests that current environmental factors remain suitable for the  
472 species, although environmental changes have occurred since, notably a tempera-  
473 ture rise of  $0.2 \text{ }^\circ\text{C/decade}$  for the period 1965 - 2004 in the 0 - 200 m water column  
474 layer (Michel et al., 2009, Valencia et al., 2019). The observed presence-absence  
475 ratio data set had higher validation indices for the fitted model compared to the

476 balanced ratio data set. Further, the corresponding binary habitat map included  
477 more areas outside the sampled area. For example, it predicted the species' presence  
478 close to the West of Ireland, where the species did indeed occur in the past (Guégen,  
479 1974) and more widely along the western Mediterranean coast where it is present  
480 nowadays (Spedicato et al., 2002).

481 The use of different validation indices (hit rate, TSS, CBI and accuracy) led to  
482 different threshold values for transforming the probability of presence into binary  
483 habitat and subsequently different habitat surface areas. Overall, the difference  
484 in the size of estimated habitat area was larger between validation indices when  
485 between data set types. Differences in predicted habitat areas according to the  
486 selected thresholding method have been reported by various authors (Nenzen and  
487 Araujo, 2011; Jimenez-Valverde and Lobo, 2007; Liu et al., 2005;2016). In our case,  
488 given the species' low prevalence, absence cell records had a large impact on the  
489 estimated threshold when maximizing using the accuracy index, leading to habitats  
490 mostly restrained to the cells where species observations had been made, hence  
491 being closer to the (only partial) observed distribution. Although its use has been  
492 recommended in the case of unbalanced prevalence (Leroy et al., 2018), the use of  
493 the CBI index led to the same effect. In contrast, using the TSS index for setting the  
494 probability threshold value involved balancing correct predictions of both presence  
495 and absence cells, and hence seemed more likely to lead to binary habitats closer to  
496 the species' potential habitats.

497 The difference between habitat areas derived using TSS optimized threshold val-  
498 ues for balanced and observed unbalanced data sets depended strongly on the region  
499 used for optimization. The largest negative difference was observed for the whole  
500 area (-53%), while it was negligible (-2%) for the NE Atlantic region and positive  
501 (27%) for the Azores. Unbalanced prevalence has been reported to artificially in-  
502 crease the TSS value (Leroy et al., 2018), which implies that the balanced data set  
503 should have provided a better understanding of blackspot seabream potential habi-  
504 tats. Contrary to this expectation it seems that in our case using a bigger data set  
505 with unbalanced presence/absence cells was more informative than using a smaller  
506 data set with a selected number of balanced presence/absence cells.



507 Several studies have shown that the use of presence-true absence data are gen-  
508 erally better than presence-only data with or without using pseudo absences, and  
509 that presence-only models generally under-estimate the species' presence in loca-  
510 tions where is has not been sampled (Dorazio, 2014; Meynard et al., 2019; Wisz  
511 and Guisan, 2009). Our study is in accordance with this general result. In addition  
512 it showed that the proportion of absences data matters. The difference between  
513 the estimated area of the blackspot seabream potential habitat obtained with the  
514 observed-ratio data set and the balanced-ratio data set for the same threshold value  
515 increased with the decreasing proportion of presence data in the observed data, from  
516 the Azores to the NE Atlantic, which in turn corresponds to decreasing stock status  
517 of blackspot seabream. In summary, the results of this study provided evidence that  
518 when using opportunistic data for SDM fitting attention needs to be paid to the  
519 effects of presence/absence data imbalance as well as the choice of validation indices  
520 to fully evaluate uncertainty of estimated habitat maps.

## 521 **Acknowledgement**

522 The study received financial support from France Filière Pêche (project DynRose)  
523 and the European Union's Horizon 2020 research and innovation programme under  
524 Grant Agreement No 773713 (PANDORA).

## 525 **References**

- 526 Afonso, P., Graça, G., Berke, G., Fontes, J., 2012. First observations on seamount  
527 habitat use of blackspot seabream (*Pagellus bogaraveo*) using acoustic telemetry. J.  
528 Exp. Mar. Bio. Ecol. 436–437, 1–10. <https://doi.org/10.1016/j.jembe.2012.08.003>
- 529 Afonso, P., McGinty, N., Graça, G., Fontes, J., Inácio, M., Totland, A., Menezes, G.,  
530 2014. Vertical migrations of a deep-sea fish and its prey. PLoS One 9. [https://doi.org](https://doi.org/10.1371/journal.pone.0097884)  
531 [/10.1371/journal.pone.0097884](https://doi.org/10.1371/journal.pone.0097884)
- 532 Albouy, C., Guilhaumon, F., Araújo, M.B., Mouillot, D., Leprieur, F., 2012. Com-  
533 bining projected changes in species richness and composition reveals climate change  
534 impacts on coastal Mediterranean fish assemblages. Glob. Chang. Biol. 18,  
535 2995–3003. <https://doi.org/10.1111/j.1365-2486.2012.02772.x>

536 Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distri-  
537 bution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.*  
538 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>

539 Burgos, C., Gil, J., Del Olmo, L.A., 2013. The Spanish blackspot seabream (*Pagellus*  
540 *bogaraveo*) fishery in the Strait of Gibraltar: Spatial distribution and fishing effort  
541 derived from a small-scale GPRS/GSM based fisheries vessel monitoring system.  
542 *Aquat. Living Resour.* 26, 399–407. <https://doi.org/10.1051/alr/2013068>

543 Clairbaux, M., Fort, J., Mathewson, P., Porter, W., Strøm, H., Grémillet, D., 2019.  
544 Climate change could overturn bird migration: Transarctic flights and high-latitude  
545 residency in a sea ice free Arctic. *Sci. Rep.* 9. [https://doi.org/10.1038/s41598-019-](https://doi.org/10.1038/s41598-019-54228-5)  
546 [54228-5](https://doi.org/10.1038/s41598-019-54228-5)

547 Crisp D.J., Southward A.J., Southward E.C., 1981. On the distribution of the inter-  
548 tidal barnacles *Chthamalus stellatus*, *Chthamalus montagui* and *Euraphia depressa*.  
549 *J Mar Biol Assoc UK* 61:359–380

550 D’Onghia, G., Maiorano, P., Carlucci, R., Capezzuto, F., Carluccio, A., Tursi, A.,  
551 Sion, L., 2012. Comparing Deep-Sea Fish Fauna between Coral and Non-Coral  
552 “Megahabitats” in the Santa Maria di Leuca Cold-Water Coral Province (Mediterranean Sea). *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0044509>

554 D’Onghia, G., Maiorano, P., Sion, L., Giove, A., Capezzuto, F., Carlucci, R., Tursi,  
555 A., 2010. Effects of deep-water coral banks on the abundance and size structure of  
556 the megafauna in the Mediterranean Sea. *Deep. Res. Part II Top. Stud. Oceanogr.*  
557 57, 397–411. <https://doi.org/10.1016/j.dsr2.2009.08.022>

558 Dardignac, J., 1988. Les pêcheries du Golfe de Gascogne, bilan des connaissances.  
559 Plouzané, IFREMER, 204 pp.

560 Desbrosses, P., 1932. Poissons de chalut - la dorade commune (*Pagellus centrodontus*  
561 Delaroche) et sa pêche. *Rev. des Trav. l’Office des Pêches Marit.* 5, 167–222.

562 Dorazio, R.M., 2014. Accounting for imperfect detection and survey bias in sta-  
563 tistical analysis of presence-only data. *Glob. Ecol. Biogeogr.* 23, 1472–1484.  
564 <https://doi.org/10.1111/geb.12216>

565 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz,  
566 J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Os-  
567 borne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S.,  
568 2013. Collinearity: A review of methods to deal with it and a simulation study eval-  
569 uating their performance. *Ecography (Cop.)*. 36, 27–46. [https://doi.org/10.1111/](https://doi.org/10.1111/j.1600-0587.2012.07348.x)  
570 [j.1600-0587.2012.07348.x](https://doi.org/10.1111/j.1600-0587.2012.07348.x)

571 Elith, J., Leathwick, J.R., 2009. Species distribution models: Ecological explanation  
572 and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697.  
573 <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

574 Erzini, K., Salgado, M., Castro, M., 2006. Dynamics of black spot sea bream (*Pag-*  
575 *ellus bogaraveo*) mean length: Evaluating the influence of life history parameters,  
576 recruitment, size selectivity and exploitation rates. *J. Appl. Ichthyol.* 22, 183–188.  
577 <https://doi.org/10.1111/j.1439-0426.2006.00702.x>

578 FAO-GFCM, 2021. Fisheries and aquaculture statistics. GFCM Catches 1970-2019  
579 (FishstatJ). In: FAO Fisheries Division [on-line]. Rome. 2021 update. [https://www.](https://www.fao.org/fishery/statistics/software/fishstatj/en)  
580 [fao.org/fishery/statistics/software/fishstatj/en](https://www.fao.org/fishery/statistics/software/fishstatj/en)

581 Fithian, W., Elith, J., Hastie, T., Keith, D.A., 2015. Bias correction in species dis-  
582 tribution models: Pooling survey and collection data for multiple species. *Methods*  
583 *Ecol. Evol.* 6, 424–438. <https://doi.org/10.1111/2041-210X.12242>

584 Fletcher, R.J., Hefley, T.J., Robertson, E.P., Zuckerberg, B., McCleery, R.A., Do-  
585 razio, R.M., 2019. A practical guide for combining data to model species distribu-  
586 tions. *Ecology* 100, 1–15. <https://doi.org/10.1002/ecy.2710>

587 Fletcher, R.J., McCleery, R.A., Greene, D.U., Tye, C.A., 2016. Integrated models  
588 that unite local and regional data reveal larger-scale environmental relationships  
589 and improve predictions of species distributions. *Landsc. Ecol.* 31, 1369–1382.  
590 <https://doi.org/10.1007/s10980-015-0327-9>

591 Francis, R.I.C.C., Clark, M.R., 2005. Sustainability issues for orange roughy fish-  
592 eries. *Bull. Mar. Sci.* 76, 337–351.

593 Gleason, H.A., 1926. The individualistic concept of plant association. *Bull. Torrey*  
594 *Bot. Club* 53, 7–26.

595 Guéguen, J. 1969. Croissance de la dorade, *Pagellus centrodontus* Delaroché. Revue  
596 des Travaux de l'Institut des Pêches Maritimes, 33: 251–264.

597 Gueguen, J., 1974. Précisions sur les migrations de la dorade rose *Pagellus Bogaraveo*  
598 (Brunnich 1768). Sci. Pêche 237, 1–10.

599 Guichet, R., Guéguen, J., Guillou, A., 1971. La pêche du merlu et de la dorade à  
600 La Rochelle, analyse des statistiques d'effort de pêche et de production des années  
601 1966, 1967 et 1968. Rev. Trav. Inst. Pêches marit. 35, 239–286.

602 Guisan, A., Thuiller, W., 2005. Predicting species distribution: Offering more than  
603 simple habitat models. Ecology Letters 8 (9), pp. 993–1009. [https://doi.org/10.1111/  
604 j.1461-0248.2005.00792.x](https://doi.org/10.1111/j.1461-0248.2005.00792.x)

605 Hareide, N.R., Garnes, G., 2001. The distribution and catch rates of deep wa-  
606 ter fish along the Mid-Atlantic ridge from 43 to 61 °N. Fish. Res. 51, 297–310.  
607 [https://doi.org/10.1016/S0165-7836\(01\)00253-3](https://doi.org/10.1016/S0165-7836(01)00253-3)

608 Helaouët, P., Beaugrand, G., 2009. Physiology, ecological niches and species distri-  
609 bution. Ecosystems, 12(8), 1235–1245. <https://doi.org/10.1007/s10021-009-9261-5>

610 Hijmans, R.J., Etten, J.V., 2012. Raster: Geographic Analysis and Modeling With  
611 Raster Data. R Package Version 2.0–12.

612 Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the  
613 ability of habitat suitability models to predict species presences. Ecol. Modell. 199,  
614 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>

615 Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harb. Symp. Quant.  
616 Biol. 22, 415–427

617 Hutchinson, G.E., 1978. An Introduction to Population Ecology. Yale University  
618 Press, New Haven.

619 ICES. 2021. Working Group on the Biology and Assessment of Deep-sea Fisheries  
620 Resources (WGDEEP). ICES Scientific Reports. 3:47. 944 pp. [http://doi.org/10.  
621 17895/ices.pub.8108](http://doi.org/10.17895/ices.pub.8108)

622 Jiménez-Valverde, A., 2021. Prevalence affects the evaluation of discrimination ca-  
623 pacity in presence-absence species distribution models. *Biodivers. Conserv.* 30,  
624 1331–1340. <https://doi.org/10.1007/s10531-021-02144-4>

625 Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of proba-  
626 bility of species presence to either-or presence-absence. *Acta Oecologica* 31, 361–369.  
627 <https://doi.org/10.1016/j.actao.2007.02.001>

628 Jiménez, L., Soberón, J., 2020. Leaving the area under the receiving operating  
629 characteristic curve behind: An evaluation method for species distribution mod-  
630 elling applications based on presence-only data. *Methods Ecol. Evol.* 2020, 1–16.  
631 <https://doi.org/10.1111/2041-210X.13479>

632 Kellner, K.F., Swihart, R.K., 2014. Accounting for imperfect detection in ecology:  
633 A quantitative review. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0111436>

634 Koslow, J. A. 1996. Energetic and life-history patterns of deep-sea benthic, ben-  
635 thopelagic and seamount-associated fish. *J. Fish Biol.* 49:54-74.

636 Laman, E.A., Rooper, C.N., Turner, K., Rooney, S., Cooper, W., Zimmermann, M.,  
637 2018. Using species distribution models to describe essential fish habitat in Alaska.  
638 *Can. J. Fish. Aquat. Sci.*

639 Leroy, B., Bellard, C., Dubos, N., Colliot, A., Vasseur, M., Courtial, C., Bakkenes,  
640 M., Canard, A., Ysnel, F., 2014. Forecasted climate and land use changes, and  
641 protected areas: The contrasting case of spiders. *Divers. Distrib.* 20, 686–697.  
642 <https://doi.org/10.1111/ddi.12191>

643 Leroy, B., Delsol, R., Hugueny, B., Meynard, C.N., Barhoumi, C., Barbet-Massin,  
644 M., Bellard, C., 2018. Without quality presence–absence data, discrimination met-  
645 rics such as TSS can be misleading measures of model performance. *J. Biogeogr.*  
646 45, 1994–2002. <https://doi.org/10.1111/jbi.13402>.

647 Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of  
648 occurrence in the prediction of species distributions. *Ecography (Cop.)*. 28, 385–393.  
649 <https://doi.org/10.1111/j.0906-7590.2005.03957.x>

650 Liu, C., Newell, G., White, M., 2016. On the selection of thresholds for predicting  
651 species occurrence with presence-only data. *Ecol. Evol.* 6, 337–348. <https://doi.org/10.1002/ece3.1878>  
652

653 Lorance, P., 2011. History and dynamics of the overexploitation of the blackspot sea  
654 bream (*Pagellus bogaraveo*) in the Bay of Biscay. *ICES J. Mar. Sci.* 68, 290–301.  
655 <https://doi.org/10.1093/icesjms/fsq072>

656 Lorance, P., F. Uiblein, and D. Latrouite. 2002. Habitat, behaviour and colour  
657 patterns of orange roughy *Hoplostethus atlanticus* (Pisces: Trachichthyidae) in the  
658 Bay of Biscay. *J. Mar. Biolog. Assoc.* 82:321-331.

659 Mateo, R., Aroca-Fernández, M.J., Gastón, A., Gómez-Rubio, V., Saura, S., García-  
660 Viñas, J.I., 2019. Looking for an optimal hierarchical approach for ecologically  
661 meaningful niche modelling. *Ecol. Modell.* 409. <https://doi.org/10.1016/j.ecolmodel.2019.108735>  
662

663 Menezes, G. M., Giacomello, E., 2013. Spatial and temporal variability of demersal  
664 fishes at Condor seamount (Northeast Atlantic). *Deep-Sea Research Part II: Topical  
665 Studies in Oceanography*, 98(PA), 101–113. <https://doi.org/10.1016/j.dsr2.2013.08.010>

666 Meynard, C.N., Leroy, B., Kaplan, D.M., 2019. Testing methods in species distribu-  
667 tion modelling using virtual species: what have we learnt and what are we missing?  
668 *Ecography (Cop.)*. 42, 2021–2036. <https://doi.org/10.1111/ecog.04385>

669 Michel, S., Vandermeirsch, F., Lorance, P., 2009. Evolution of upper layer tem-  
670 perature in the bay of biscay during the last 40 years. *Aquat. Living Resour.* 22,  
671 447–461. <https://doi.org/10.1051/alr/2009054>

672 Montgomery, D.C., 2005. *Design and Analysis of Experiments*. Wiley

673 Morato, T., Solà, E., Grós, M.P., Menezes, G., 2001. Feeding habits of two congener  
674 species of seabreams, *Pagellus bogaraveo* and *Pagellus acarne*, off the Azores (north-  
675 eastern Atlantic) during spring of 1996 and 1997. *Bull. Mar. Sci.* 69, 1073–1087.

676 Mytilineou, C., Politou, C.Y., Papaconstantinou, C., Kavadas, S., D’Onghia, G.,  
677 Sion, L., 2005. Deep-water fish fauna in the Eastern Ionian Sea. *Belgian J. Zool.*  
678 135, 229–233.

679 Mytilineou, C., Smith, C.J., Anastasopoulou, A., Papadopoulou, K.N., Christidis,  
680 G., Bekas, P., Kavadas, S., Dokos, J., 2014. New cold-water coral occurrences in the  
681 Eastern Ionian Sea: Results from experimental long line fishing. Deep. Res. Part  
682 II Top. Stud. Oceanogr. 99, 146–157. <https://doi.org/10.1016/j.dsr2.2013.07.007>

683 Nenzén, H.K., Araújo, M.B., 2011. Choice of threshold alters projections of species  
684 range shifts under climate change. Ecol. Modell. 222, 3346–3354. <https://doi.org/10.1016/j.ecolmodel.2011.07.011>

686 Olivier, R., 1928. Poissons de chalut, la dorade (*Pagellus centrodontus*). Rev. des  
687 Trav. l'Office des Pêches Marit. 1, 5–32.

688 Pecchi, M., Marchi, M., Burton, V., Giannetti, F., Moriondo, M., Bernetti, I., Bindi,  
689 M., Chirici, G., 2019. Species distribution modelling to support forest management.  
690 A literature review. Ecol. Modell. 411, 108817. [https://doi.org/10.1016/j.ecolmod](https://doi.org/10.1016/j.ecolmodel.2019.108817)  
691 [el.2019.108817](https://doi.org/10.1016/j.ecolmodel.2019.108817)

692 Pinho, M., Diogo, H., Carvalho, J., J., P.G., 2014. Harvesting juveniles of blackspot  
693 sea bream (*Pagellus bogaraveo*) in the Azores (Northeast Atlantic): biological im-  
694 plications, management, and life cycle considerations. ICES J. Mar. Sci. 71,  
695 2448–2456. <https://doi.org/10.4135/9781412953924.n678>

696 Rincón, L., Castro, P.L., Álvarez, B., Hernández, M.D., Álvarez, A., Claret, A.,  
697 Guerrero, L., Ginés, R., 2016. Differences in proximal and fatty acid profiles, sen-  
698 sory characteristics, texture, colour and muscle cellularity between wild and farmed  
699 blackspot seabream (*Pagellus bogaraveo*). Aquaculture 451, 195–204. [https://doi.](https://doi.org/10.1016/j.aquaculture.2015.09.016)  
700 [org/10.1016/j.aquaculture.2015.09.016](https://doi.org/10.1016/j.aquaculture.2015.09.016)

701 Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E., Lundquist, C.J.,  
702 2017. A systematic review of marine-based Species Distribution Models (SDMs)  
703 with recommendations for best practice. Front. Mar. Sci. 4, 1–11. [https://doi.org](https://doi.org/10.3389/fmars.2017.00421)  
704 [/10.3389/fmars.2017.00421](https://doi.org/10.3389/fmars.2017.00421)

705 Santos, R., W. Medeiros-Leal, A. Novoa-Pabon, H. Silva, and M. Pinho. 2021.  
706 Demersal fish assemblages on seamounts exploited by fishing in the Azores (NE  
707 Atlantic). J. Appl. Ichthyol. 37:198-215.

708 Sanz-Fernández, V., Gutiérrez-Estrada, J.C., Pulido-Calvo, I., Gil-Herrera, J., Ben-  
709 choucha, S., el Arraf, S., 2019. Environment or catches? Assessment of the decline  
710 in blackspot seabream (*Pagellus bogaraveo*) abundance in the Strait of Gibraltar. *J.*  
711 *Mar. Syst.* 190, 15–24. <https://doi.org/10.1016/j.jmarsys.2018.08.005>

712 Schickele, A., Leroy, B., Beaugrand, G., Goberville, E., Hattab, T., Francour, P.,  
713 Raybaud, V., 2020. Modelling European small pelagic fish distribution: Method-  
714 ological insights. *Ecol. Modell.* 416, 108902. [https://doi.org/10.1016/j.ecolmodel.](https://doi.org/10.1016/j.ecolmodel.2019.108902)  
715 [2019.108902](https://doi.org/10.1016/j.ecolmodel.2019.108902)

716 Silva, H. M., Pinho, M. R. (2007). Small-Scale Fishing on Seamounts. In T. J.  
717 Pitcher, T. Morato, P. J. B. Hart, M. R. Clark, N. Haggan, and R. S. Santos  
718 (Eds.), *Seamounts: Ecology, Fisheries and Conservation* (pp. 335–360). Blackwell  
719 Publishing.

720 Soberón, J., Nakamura, M., 2009. Niches and distributional areas: Concepts, meth-  
721 ods, and assumptions. *Proceedings of the National Academy of Sciences of the*  
722 *United States of America*, 106 (SUPPL. 2), 19644–19650. [https://doi.org/10.1073/](https://doi.org/10.1073/pnas.0901637106)  
723 [pnas.0901637106](https://doi.org/10.1073/pnas.0901637106)

724 Spedicato, M.T., Greco, S., Sophronidis, K., Lembo, G., Giordano, D., Argyri, A.,  
725 2002. Geographical distribution, abundance and some population characteristics of  
726 the species of the genus *Pagellus* (Osteichthyes: Perciformes) in different areas of the  
727 Mediterranean. *Sci. Mar.* 66, 65–82. <https://doi.org/10.3989/scimar.2002.66s265>

728 Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD - A platform  
729 for ensemble forecasting of species distributions. *Ecography (Cop.)*. 32, 369–373.  
730 <https://doi.org/10.1111/j.1600-0587.2008.05742.x>

731 Thuillier, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., 2016. Biomod2:  
732 Ensemble platform for species distribution modeling. *R Packag. version 3.3-7*.

733 Valanis, D. V., Pierce, J.P., Zuur, A.F., Pali Alexis, A., Saveliev, A., Katara, I.,  
734 Wang, J., 2008. Modelling of essential fish habitat based on remote sensing, spatial  
735 analysis and GIS. *Hydrobiologia* 612, 5–20.

736 Valencia, V., Fontán, A., Goikoetxea, N., Chifflet, M., González, M., López, A.,  
737 2019. Long-term evolution of the stratification, winter mixing and  $\Theta$ -S signature



738 of upper water masses in the southeastern Bay of Biscay. *Cont. Shelf Res.* 181,  
739 124–134. <https://doi.org/10.1016/j.csr.2019.05.010>

740 Welsh, A.H., Lindenmayer, D.B., Donnelly, C.F., 2013. Fitting and Interpreting  
741 Occupancy Models. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0052015>

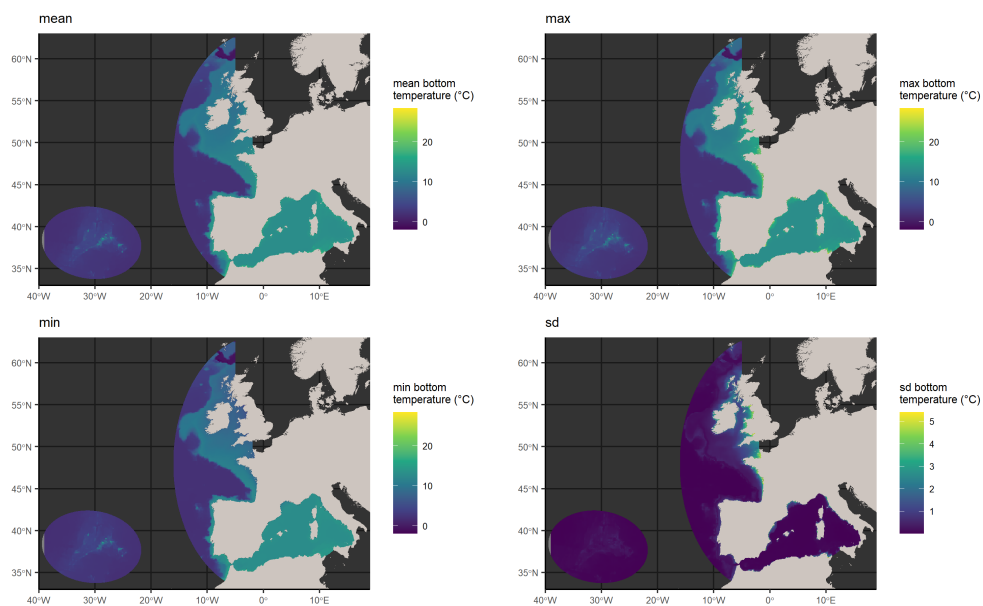
742 Wiens, J.J., Sukumaran, J., Pyron, R.A., Brown, R.M., 2009. Evolutionary and  
743 biogeographic origins of high tropical diversity in Old World frogs (Ranidae). *Evo-*  
744 *lution*, 63, 1217-1231.

745 Wisz, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence  
746 species distribution models and their predictions? An information-theoretic ap-  
747 proach based on simulated data. *BMC Ecol.* 9, 1–13. [https://doi.org/10.1186/1472-](https://doi.org/10.1186/1472-6785-9-8)  
748 [6785-9-8](https://doi.org/10.1186/1472-6785-9-8)

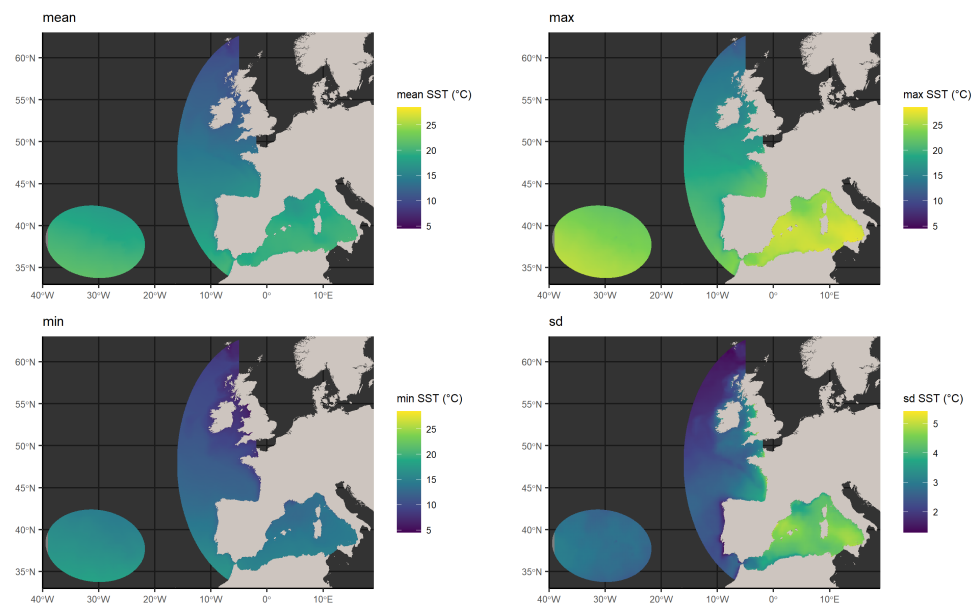
749 Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos,  
750 G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J.,  
751 Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-  
752 Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A  
753 standard protocol for reporting species distribution models. *Ecography (Cop.)*. 43,  
754 1261–1277. <https://doi.org/10.1111/ecog.04960>

755 **Supplementary Material**

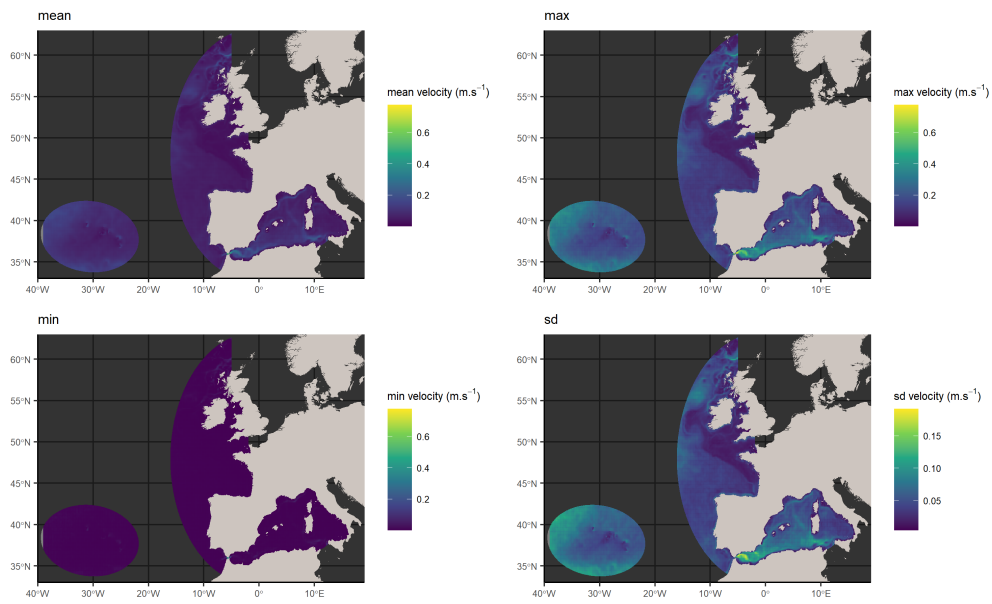
756 *Supplementary Material 1: Environmental variables extracted*



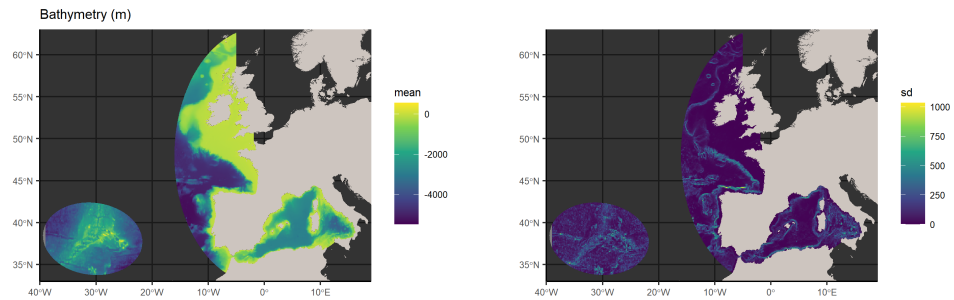
757 Sup. Mat. 1.1: Annual mean, maximum, minimum, and standard deviation of the bottom  
758 temperature over the blackspot sea bream main area of distribution between January 1994 and  
759 December 2018 extracted from Copernicus.



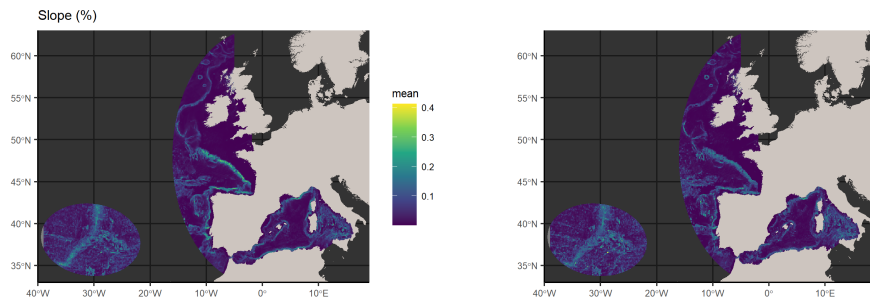
760 Sup. Mat. 1.2: Annual mean, maximum, minimum, and standard deviation of the sea surface  
761 temperature (SST) over the blackspot sea bream main area of distribution between January 1994  
762 and December 2018 extracted from Copernicus.



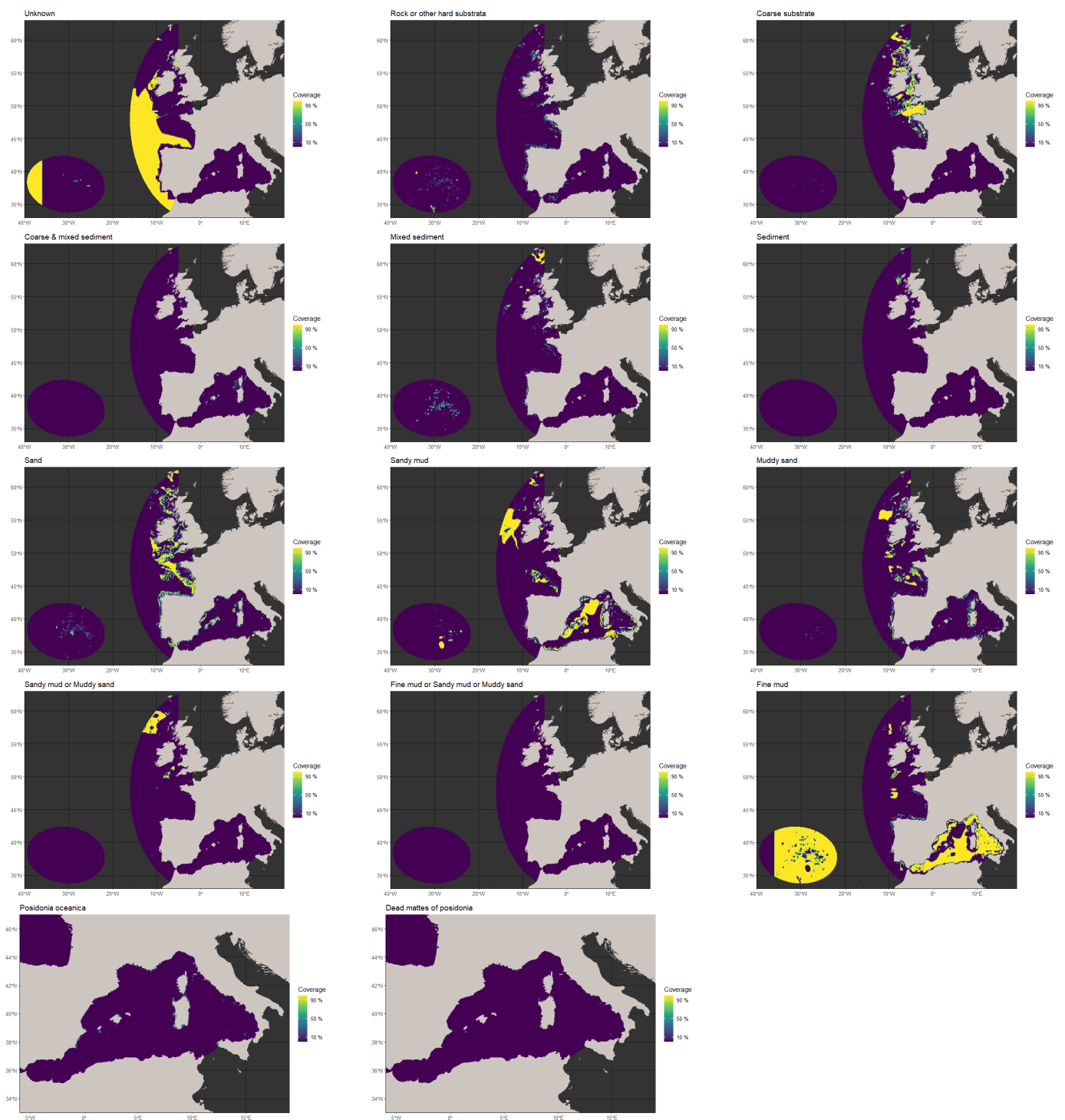
763 Sup. Mat. 1.3: Annual mean, maximum, minimum, and standard deviation of the absolute  
 764 currents velocity over the blackspot sea bream main area of distribution between January 1994 and  
 765 December 2018 extracted from Copernicus.



766 Sup. Mat. 1.4: Mean and standard deviation of the bathymetry over the blackspot seabream  
 767 main area of distribution extracted from GEBCO.

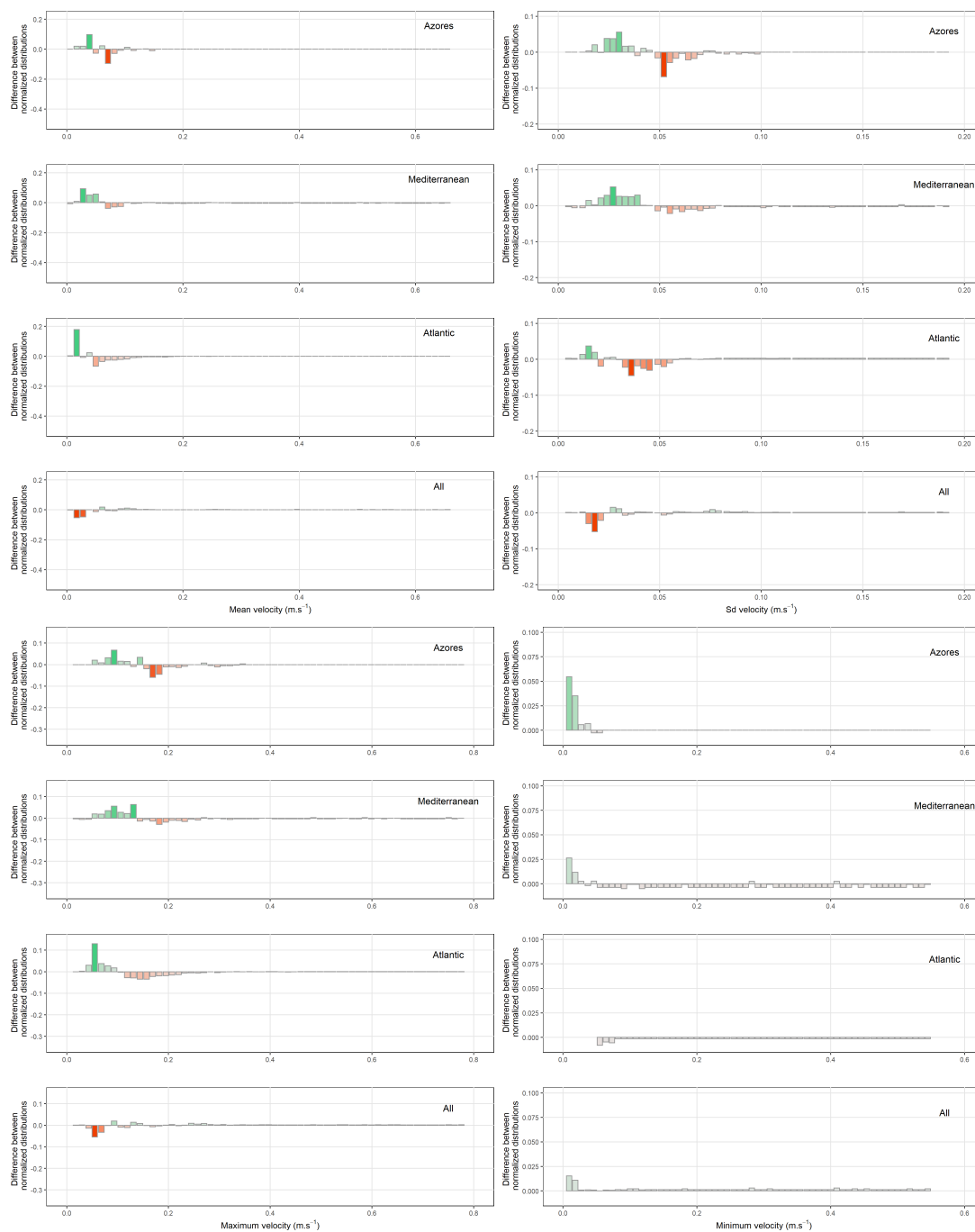


768 Sup. Mat. 1.5: Mean and standard deviation of the slope over the blackspot seabream main  
 769 area of distribution extracted from GEBCO.

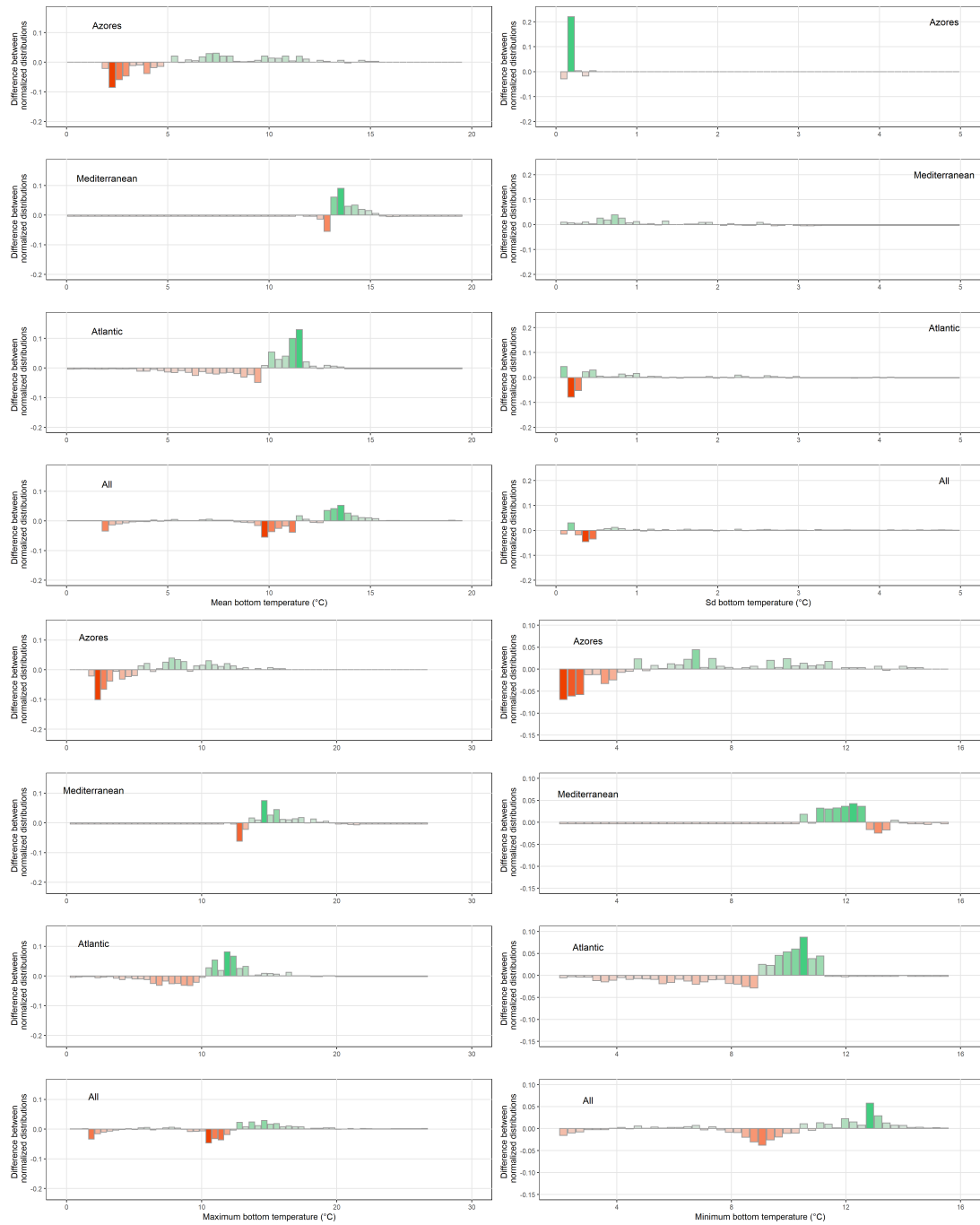


770 Sup. Mat. 1.6: Percentage of cell occupied by each of the 14 Emodnet sea bottom categories:  
 771 unknown, rock or other hard substrata, coarse substrate, coarse and mixed sediment, mixed sed-  
 772 iment, sediment, sand, sandy mud, muddy sand, sandy mud or muddy sand, fine mud or sandy  
 773 mud or muddy sand, fine mud, *Posidonia oceanica*, dead mattes of *Posidonia oceanica*. Each cell  
 774 was attributed the value corresponding to the most present category.

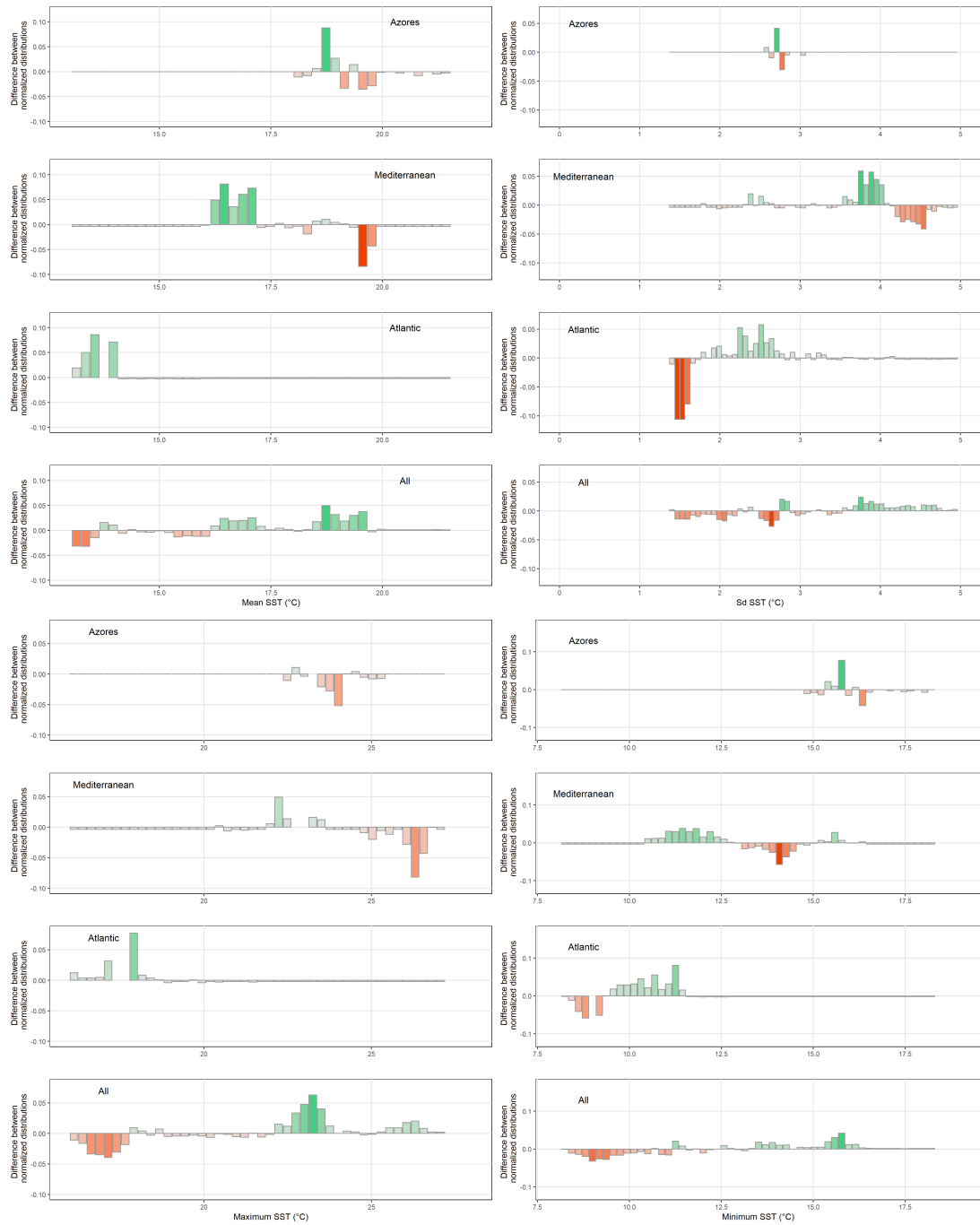
775 *Supplementary Material 2: Data exploration*



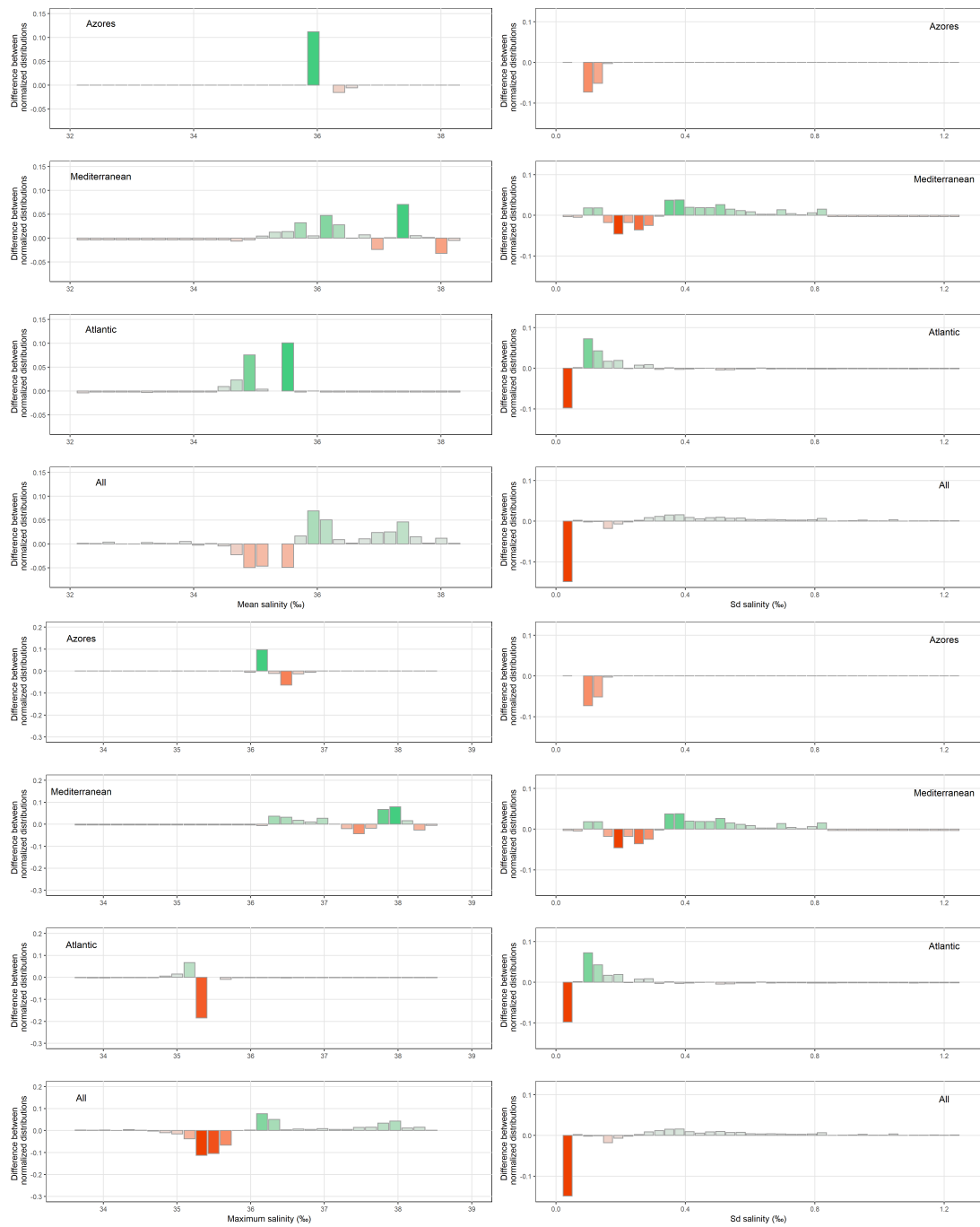
776 Sup. Mat. 2.1: Comparison of the distributions in mean, sd, maximum and minimum annual  
 777 currents velocity of presence records over sampling cells in the whole area and in Azorean region,  
 778 Mediterranean Sea, and Atlantic shelf.



779 Sup. Mat. 2.2: Comparison of the distributions in mean, sd, maximum and minimum annual  
 780 bottom temperature of presence records over sampling cells in the whole area and in Azorean  
 781 region, Mediterranean Sea, and Atlantic shelf.

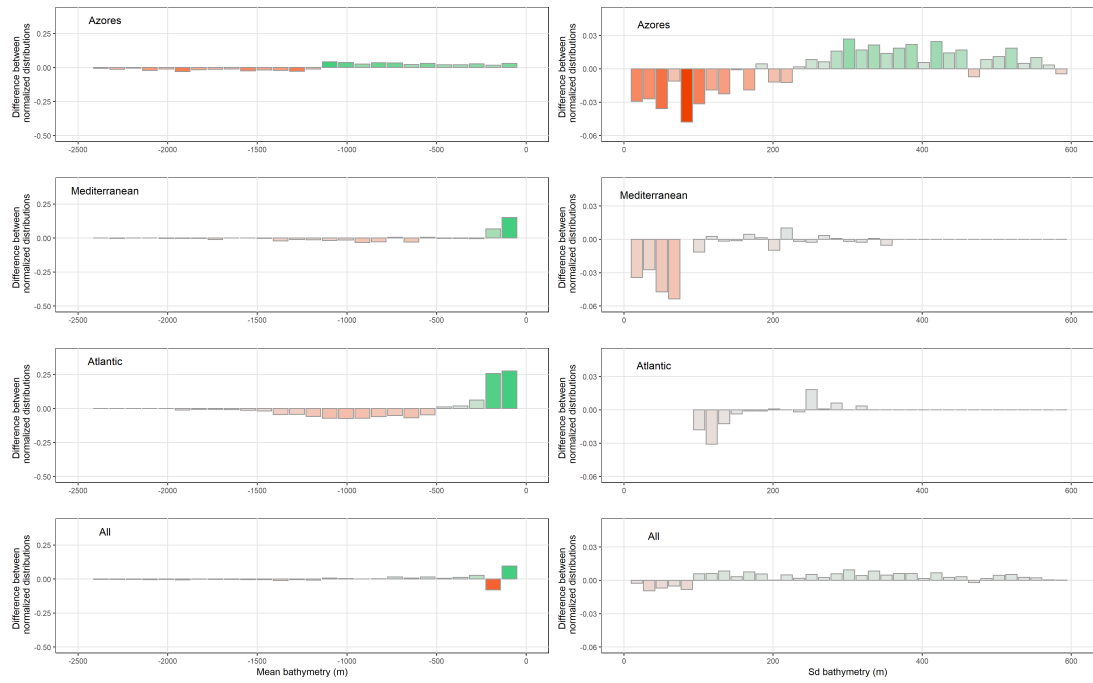


782 Sup. Mat. 2.3: Comparison of the distributions in mean, sd, maximum and minimum annual  
 783 sea surface temperature (SST) of presence records over sampling cells in the whole area and by  
 784 region (Azores, Mediterranean Sea, and Atlantic shelf).



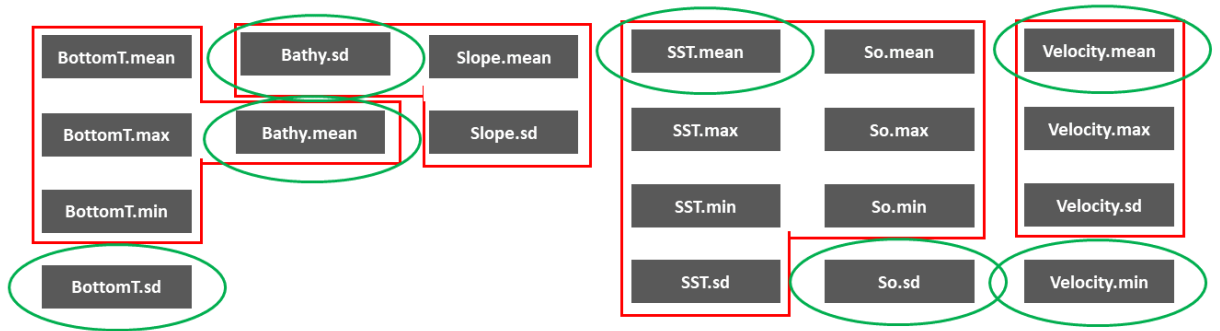
785 Sup. Mat. 2.4: Comparison of the distributions in mean, sd, maximum and minimum annual  
 786 salinity of presence records over sampling cells in the whole area and in Azorean region,  
 787 Mediterranean Sea, and Atlantic shelf.





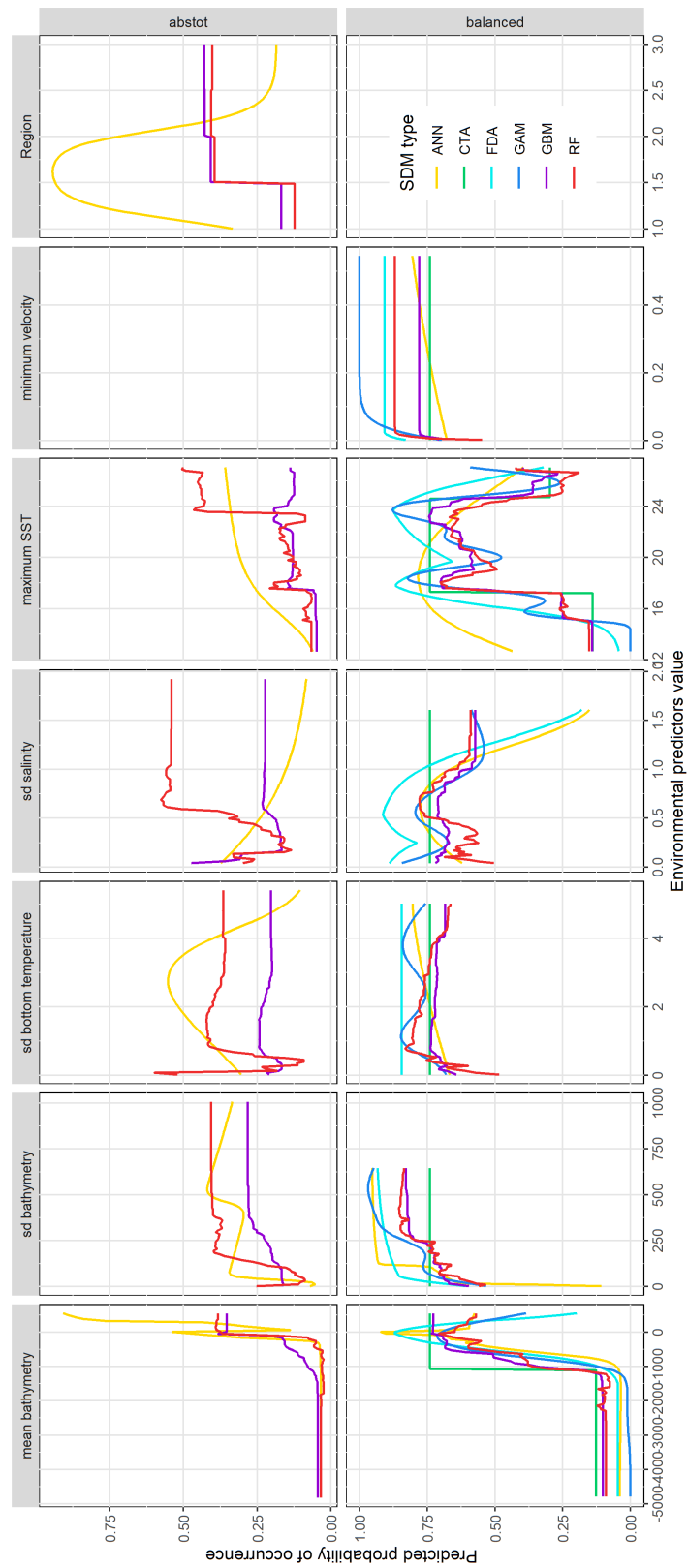
788 Sup. Mat. 2.5: Comparison of the distributions in mean and sd bathymetry of presence records  
 789 over sampling cells in the whole area and in Azorean region, Mediterranean Sea, and Atlantic shelf.

790 *Supplementary Material 3: Groups of correlated environmental variables*



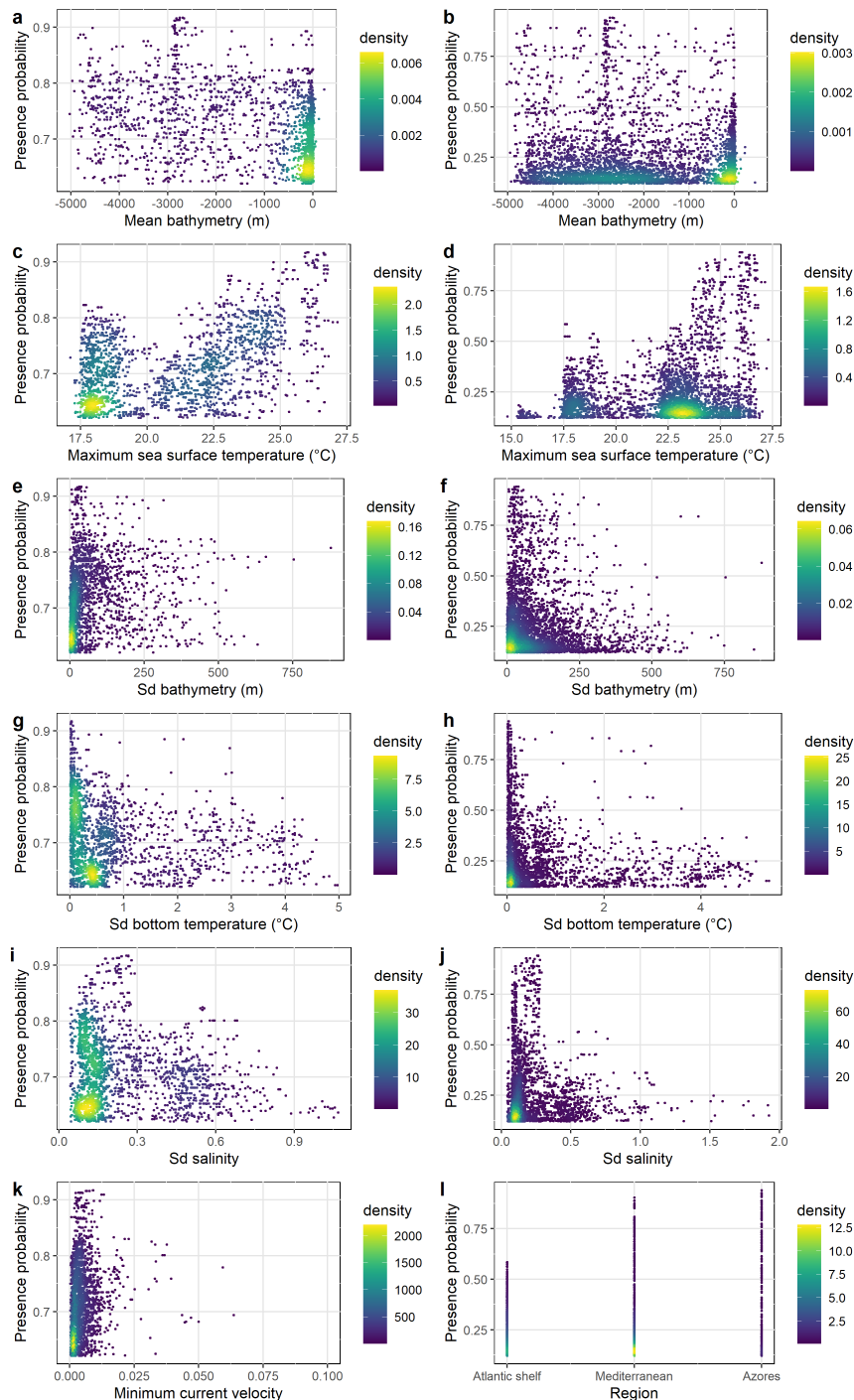
791 Sup. Mat. 3.1: Groups of correlated variables (Pearson's correlation,  $r > 0.7$ ) in red, and vari-  
 792 able selected among each group, in green, selected according the percentage of explained variance.  
 793 BottomT is bottom temperature, Bathy is bathymetry, SST sea surface temperature, So salinity  
 794 and Velocity surface current velocity.

795 *Supplementary Material 4: Response curves of the various selected SDMs used in the ensemble model procedure using 6 environmental*  
 796 *variables with the observed- (top) and balanced-ratio (bottom) data sets.*





797 *Supplementary Material 5: Distribution of presence cells of the blasckspot sea*  
 798 *breem in the model domain of distribution (based on the presence threshold obtained*  
 799 *with TSS indices: presence respectively for a presence probability  $p > 0.62$  for the*  
 800 *balanced-ratio data set and  $p > 0.12$  for the observed-ratio data set) according to*  
 801 *the predictors values when using 6 predictors with the balanced- (a, c, e, g, i, k) and*  
 802 *observed-ratio (b, d, f, h, j, l) data sets. High densities of values appear yellow on*  
 803 *the scatter plots.*

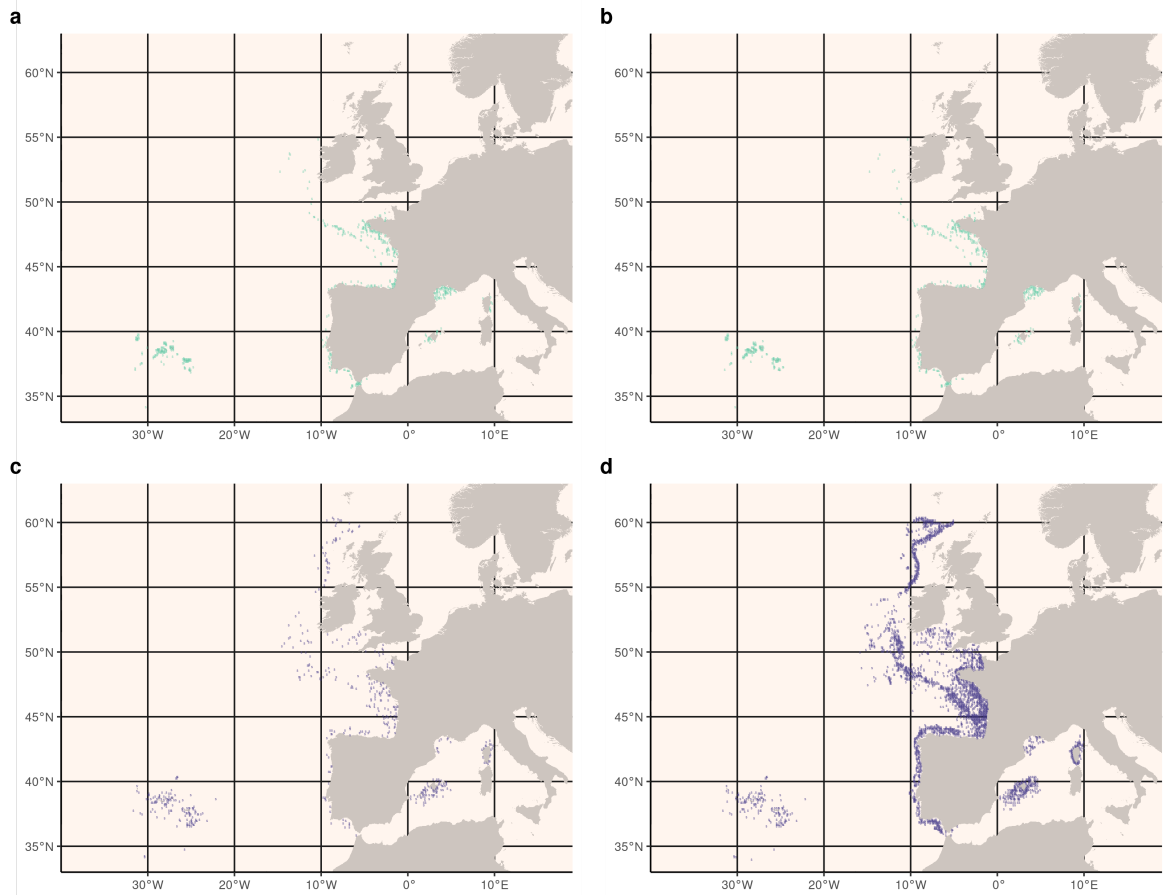




804 *Supplementary Material 6: Summary of the 16 ensemble species distribution models implemented in this study with 8 different*  
805 *numbers of predictors and two different kinds of data sets (balanced- / observed-ratio): predictors used and their relative importance*  
806 *(proportion of variance explained), characteristics of the data set used, used SDMs and overall performance.*

	Data set	2 predictors	3 predictors	4 predictors	5 predictor	6 predictors	7 predictors	8 predictors	9 predictors
<b>Environmental variables' importance</b>									
Mean bathymetry	balanced	64%	69%	60%	59%	55%	55 %	53%	58%
	observed	50%	49 %	50%	37%	30%	31%	26%	27 %
Max SST	balanced	36%	22%	23 %	28%	28%	22%	28%	22%
	observed	50%	34%	18%	25%	31%	25%	27%	26%
Sd bathymetry	balanced	-	9%	12 %	7%	9%	12%	9%	10%
	observed	-	17%	18%	20%	18%	16%	12%	11%
Sd bottom temperature	balanced	-	-	5%	7%	3%	2%	3%	3%
	observed	-	-	-	8%	8%	9%	7 %	8%
Sd salinity	balanced	-	-	-	2 %	3%	3%	3%	2%
	observed	-	-	-	-	5%	6 %	5 %	5%
Min current velocity	balanced	-	-	-	-	2%	2%	2%	1%
	observed	-	-	-	-	-	3%	4%	2%
Mean current velocity	balanced	-	-	-	-	-	2 %	1%	1%
	observed	-	-	-	-	-	-	-	2 %
Sea bottom type	balanced	-	-	-	-	-	-	1%	1%
	observed	-	-	-	-	-	-	1%	1%
Region	balanced	-	-	-	-	-	-	-	1%
	observed	-	-	11%	13 %	9%	10%	18%	17%
<b>Occurrence data set</b>									
Data set size	balanced	16	322	322	744	940	950	1072	1152
	observed	474	1510	1696	2620	3280	3357	3752	4163
Number of presences	balanced	8	161	161	372	470	475	536	576
	observed	8	161	161	372	470	475	536	576
Number of absences	balanced	8	161	161	372	470	475	536	576
	observed	466	1349	1535	2248	2810	2882	3216	3587
<b>Models' selection and performance</b>									
	balanced	CTA, FDA,	ANN, CTA,	ANN, CTA, FDA,	ANN, CTA,	ANN, CTA,	ANN, CTA,	ANN, CTA,	ANN, CTA,
		GAM, GBM,	FDA, GAM,	GAM, GLM,	FDA, GAM,	FDA, GAM,	FDA, GAM,	FDA, GAM,	FDA, GBM,
		RF	GBM, RF	GBM RF, SRE	GBM, RF	GBM, RF	GBM, RF	GBM, RF	RF
Selected models (TSS > 0.5)	observed	ANN, FDA,	ANN, CTA,	ANN, CTA,	ANN, GBM,	ANN, GBM,	ANN, GBM,	ANN, GBM,	ANN, GBM,
		GAM, GBM,	FDA, GAM,	FDA, GBM,	GLM, RF	RF	RF	RF	RF
	balanced	0.711 ± 0.010	0.767 ± 0.004	0.712 ± 0.011	0.711 ± 0.006	0.716 ± 0.003	0.730 ± 0.001	0.7155 ± 0.014	0.725 ± 0.025
True Skill Statistic	observed	0.996 ± 0.003	0.791 ± 0.013	0.826 ± 0.013	0.867 ± 0.033	0.916 ± 0.030	0.895 ± 0.039	0.904 ± 0.043	0.894 ± 0.053

807 *Supplementary Material 7: Maps of the presence (a,b, green) and absence (c,d,*  
808 *blue) data cells used for calibrating eSDMs with the balanced-ratio (a,c) and the*  
809 *observed-ratio (b,d) of presence and absence.*





810 *Supplementary Material 8: Maximized index values when comparing the blackspot*  
811 *seabream eSDM habitat estimates for best model (6 predictors) to occurrence data and*  
812 *corresponding probability threshold values for presence/absence for building binary*  
813 *habitat maps.*

Validation index	Region	Data set	Maximized index	Optimised threshold
TSS	all	balanced	0.45	0.62
	all	observed	0.57	0.12
	Azores	balanced	0.57	0.12
	Azores	observed	0.56	0.14
	Atlantic	balanced	0.45	0.46
	Atlantic	observed	0.53	0.09
	Mediterranean	balanced	0.54	0.62
	Mediterranean	observed	0.62	0.25
accuracy	all	balanced	0.90	0.77
	all	observed	0.91	0.33
	Azores	balanced	0.78	0.25
	Azores	observed	0.79	0.23
	Atlantic	balanced	0.93	0.80
	Atlantic	observed	0.93	0.34
	Mediterranean	balanced	0.82	0.67
	Mediterranean	observed	0.84	0.3
CBI	all	balanced	0.98	1
	all	observed	1	0.98
	Azores	balanced	0.75	0.96
	Azores	observed	0.97	0.82
	Atlantic	balanced	0.97	0.79
	Atlantic	observed	0.65	0.42
	Mediterranean	balanced	0.89	1
	Mediterranean	observed	0.93	0.98
hit rate	all	balanced	1	0.04
	all	observed	1	0.01
	Azores	balanced	1	0.03
	Azores	observed	1	0.02
	Atlantic	balanced	1	0.04
	Atlantic	observed	1	0.01
	Mediterranean	balanced	1	0.03
	Mediterranean	observed	1	0.01