

1 **Running head:** Absences marine megafauna niche modelling

2

3

4 **To be or not to be: the role of absences in niche modelling for highly mo-**  
5 **bile species in dynamic marine environments**

6

7

8 Fernandez M<sup>1,2,\*</sup>, Sillero N<sup>3</sup>, Yesson C<sup>4</sup>

9

10 1. MARE - Marine and Environmental Sciences Centre, Agência Regional para o

11 Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI), Funchal, Portugal

12 2. cE3c/Azorean Biodiversity Group, Departamento de Biologia, Faculdade de Ciências e

13 Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal

14 3. CICGE - Centro de Investigação em Ciências Geo-Espaciais, Alameda do Monte da

15 Virgem, 4430-146 Vila Nova de Gaia, Portugal

16 4. Zoological Society of London, Institute of Zoology, London, UK

17

18

19 \* Corresponding author

20

21 E-mail:

22 marc.fern@gmail.com

23

24 Species distribution models are valuable tools for conservation management. However,  
25 there remain challenges in developing and interpreting these models in the marine environ-  
26 ment, such as the nature of the species used for the modelling process. When working with  
27 mobile species in dynamic environments, lack of observation is usually interpreted as an  
28 observation of absence, which can result in the introduction of biases by methodological  
29 (false) absences. Here, we explore the role of absences when modelling marine megafauna  
30 distributions. To better understand how the use of absences (or equivalent) affects the niche  
31 modelling algorithms, we used a set of 20 virtual species with different relations to the  
32 habitat (generalist static, specialist static, generalist dynamic and specialist dynamic) with  
33 different encounter rates. We tested six different modelling techniques divided into three  
34 distinct groups: presence-only, presence-background and presence-absence. We compared  
35 the outputs of the models using traditional validation metrics and overlap metrics in the  
36 geographical and environmental spaces. Algorithms characterized the ecological niche for  
37 the simulated species differently. Approaches using background data generally outper-  
38 formed the other methods, suggesting that the non-observation of a species in a given loca-  
39 tion and time should not be considered as an absence. A very intense (practically unreal-  
40 istic) sampling schema would be required to obtain a genuine unbiased absence when  
41 working with these species and habitats. For highly mobile species, a precautionary ap-  
42 proach would be to consider the non-observation of a species as part of the background (a  
43 sample of the conditions available in the study area) rather than an absence. A good starting  
44 point would be to use presence-background models, complemented with presence-absence  
45 and/or presence-only models, comparing outputs from the different algorithms tested in the  
46 geographic and environmental space. Improving model performance for highly mobile ma-  
47 rine species should lead to better-informed decision making for conservation.

48 **Keywords:** Absences, Ecological Niche Modelling, cetaceans, marine megafauna, mobile  
49 species, Species Distribution Models.

50

## 51 **INTRODUCTION**

52 A good understanding of species distributions is crucial to supporting conservation  
53 management. Improving access to technology and species occurrences data has expanded  
54 the use of modelling procedures to provide distribution estimates, using various techniques  
55 commonly termed Ecological Niche Models (ENMs) or Species Distribution Models  
56 (SDMs) (Sillero 2011; Sillero et al 2021). Over the past 20 years, more than 6000 studies  
57 have used these tools, using a wide variety of methods and protocols. This wide variation of  
58 applications has led to a set of best practices for guiding the building of the models and  
59 evaluating the models' adequacy to feed biodiversity assessments (Araujo et al., 2019;  
60 Sillero et al., 2021). One of the most discussed topics is the relative performance of  
61 different modelling algorithms in calculating distributional estimates. Previous studies tried  
62 to rate the algorithms available to select the best performing' methods (e.g. Brotons et al.,  
63 2004, Elith et al., 2006). There are three basic modelling approaches to estimate  
64 distributions available according to the nature of the characteristics of the occurrence data  
65 (Sillero et al 2021): (1) using presence-only data; (2) using presence-absence data when  
66 available (or 'pseudo-absences', when not available) and (3) using presence data and a  
67 sample of background data. According to Phillips et al. (2009), background data is a sample  
68 from the study area as a whole, used to characterize the environmental conditions across the  
69 study region, including the sites where the species could also be sighted. However, a  
70 "pseudo-absence" record intends to mimic absence data, and it might include locations that

71 the species does or could occupy, contrary to the methodology's purpose. In fact, pseudo-  
72 absences can be considered a biased sample of the background data (Franklin 2010).  
73 Modelling algorithms using presences and true-absences calculate the actual probability of  
74 the species' presence and estimate the actual distributional area, distinguishing between  
75 occupied and non-occupied areas by the species, something similar to the realized niche  
76 (RN). Conversely, background or presence-only data methods provide information on how  
77 to classify sites by their similarity to known occupied regions (Sillero et al 2021).  
78 Consequently, background or pseudo-absence data methods will predict an area between  
79 the actual occupied area and the abiotic niche (Jiménez-Valverde et al 2008; Soberon &  
80 Nakamura 2009), somehow closer to the fundamental niche concept (FN).

81 Soberon & Peterson 2005 define the RN as the set of locations where abiotic conditions are  
82 suitable for positive population growth, and the required biotic conditions (e.g. mutualists,  
83 competitors, predators...) will lead to positive fitness. On the other side, the FN represents  
84 the conditions where the conditions are suitable for the species. In the case of cetaceans,  
85 one might argue that due to their mobility and dispersion capacity, all the suitable areas will  
86 be accessible. However, the optimal biotic conditions for the species will be only present in  
87 a small range of the suitable area. Therefore, a good amount of occurrence points in the  
88 areas with suitable abiotic conditions would be needed to estimate the FN, while the RN  
89 calculation would require much more information regarding the biotic interactions.

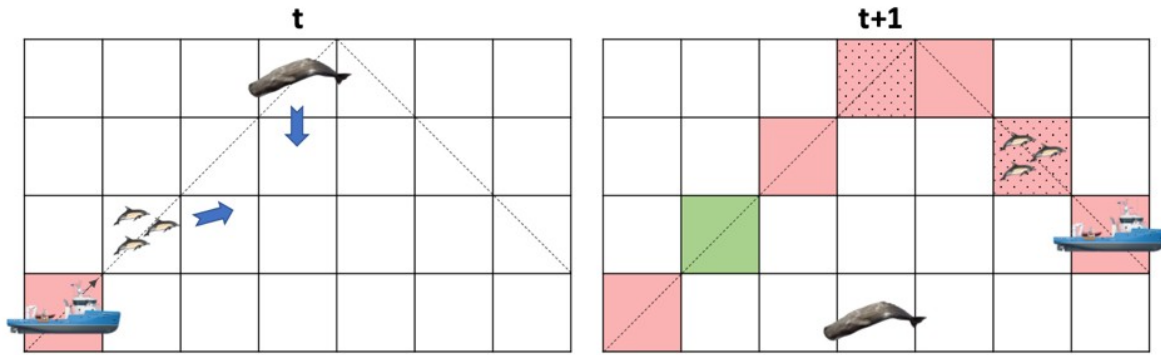
90 The best choice to produce accurate distribution estimates for current biodiversity  
91 assessments would appear to be presence-absence algorithms. However, this statement is  
92 more complex than it looks: it is necessary to consider the nature of the species data used in  
93 the modelling processes. While the concept of presence is relatively straightforward (a

94 place where the organism has been directly observed), absence is a more complex concept.  
95 They can be classified into three types: environmental, contingent and methodological  
96 (Lobo et al., 2010). The first two kinds of absences are related to natural processes shaping  
97 species distributions (Soberon & Nakamura 2009) and linked with the available  
98 environmental conditions. Environmental absences are more probable in those localities  
99 showing environmental conditions very different from the environmental universe defined  
100 by the presence localities (e.g. a polar habitat for a tropical species).

101 In contrast, contingent absences correspond to environmentally suitable areas not occupied  
102 due to biotic, historical and dispersal reasons (Lobo et al., 2010). However, methodological  
103 absences (created by biased surveys) are a consequence of the very nature of the  
104 biodiversity distribution information, which frequently is incomplete and biased. It can be  
105 challenging to interpret the absence of observation as an observation of absence (e.g.  
106 Dennis et al., 1999, Graham et al., 2004, Soberon & Peterson, 2004, Hortal et al., 2008,  
107 Lobo et al., 2007, Peterson et al., 2011, Kramer-Schadt et al., 2013, Boira et al., 2014).  
108 Methodological absences constitute the most important source of uncertainty for analysing  
109 patterns and processes underlying the geographic distribution of biodiversity (the so-called  
110 Wallacean shortfall; Whittaker et al., 2005). Therefore, while environmental absences will  
111 be prevalent in areas less environmentally suitable for the species, methodological absences  
112 can be found everywhere and are generally higher in suitable environmental locations close  
113 to the recorded presences (Lobo et al., 2010) because survey effort is naturally biased  
114 towards areas of expected presence. Other problems can arise even in a stable terrestrial  
115 ecosystem with a very intense sampling program (where confidence in ‘absences’ is high).  
116 Maybe the species is not present because the area is outside the dispersal range, or the

117 species was not present at the time of sampling due to temporary occupancy (migration) or  
118 temporal inactivity (diurnal species, nocturnal species, estivation, hibernation), or present  
119 nearby but not observed (Soberón and Peterson 2005).

120 When looking at the marine environment, these issues become even more complex. Marine  
121 environments are more dynamic than terrestrial ecosystems (Fernandez et al., 2017). The  
122 environmental suitability can rapidly change in a specific location (e.g. due to changing  
123 ocean currents), which intrinsically implies that species might move to track favourable  
124 conditions. Marine physical processes (such as eddies, upwellings and surface currents) are  
125 highly dynamic, requiring a careful selection of the predictor variables, and many times  
126 resulting in temporal or spatial lags between physical and biological processes (Redfern et  
127 al. 2006). Moreover, some marine species (such as cetaceans) are highly mobile, with daily  
128 ranges of 100+ km (Leatherwood & Ljungblad 1979; Perrin et al., 1979; Scott et al., 1993).  
129 The environmental variability and high motility of species create difficulties for obtaining  
130 complete and unbiased datasets. Therefore, methodological absences are often present in  
131 the dataset, even if assuming a perfect detectability scenario (improbable in most marine  
132 settings). For example, an area sampled at a specific time  $t+1$  might be noted as an absence;  
133 however, the species could be present at that exact location at time  $t$  (Fig. 1). While this set  
134 of environmental conditions would be noted as not favourable for the species, they could be  
135 highly favourable but simply not visited when the species was present. It is essential to  
136 highlight that modelling algorithms use surface sightings of marine species as presences.  
137 However, the species might be present at that location but simply below the surface. In  
138 such cases, the species is present but noted as absent.



139

140 **Figure 1.** Example of a simulated transect survey for cetaceans, where the boat follows a  
 141 pre-designed transect. Two potential encounters are represented (symbolized as dolphins  
 142 and a whale). As time changes from  $t$  to  $t+1$ , the animals and the boat moves, mimicking a  
 143 hypothetical situation in a real-world scenario. Red cells represent true-absence, red pointed  
 144 cells represent methodological absences and green cells observed presences.

145

146 Occurrence datasets for cetacean habitat modelling studies may come from carefully  
 147 designed, standardised field studies, including ship, aerial and acoustic surveys (Redfern et  
 148 al., 2006). Those surveys are generally based on transect distance sampling methods  
 149 (Buckland et al., 2001) to estimate quantitative abundance. Surveys are typically designed  
 150 to ensure equal sampling probabilities throughout the study area. Moreover, there are tools  
 151 to correct the errors related to the perception bias, survey conditions (e.g. sea state,  
 152 visibility), and availability bias (proportion of time spent by the animals at the surface). The  
 153 use of detectability functions, independent observers and dual-platform methods (two  
 154 observers searching for and counting animals simultaneously, ensuring they do not cue  
 155 each other on the locations of the animals) are generally used to minimize biases (Buckland

156 et al., 2001). These correction methods analyse the observer's probability of detecting the  
157 species if it was present at the sampling moment; however, generally, they do not consider  
158 the habitat dynamism or the species motility. This is not a problem for species with a low  
159 dynamism, but it is for many oceanic species, such as cetaceans. Using absences in  
160 modelling algorithms may not forecast the potential distribution of oceanic species but the  
161 sampling effort of the surveys. Therefore, datasets collected following these procedures  
162 could present a high rate of methodological absences. Recently, Glennie et al. 2020 built a  
163 theoretical framework to incorporate the mobility factor in the detectability indices;  
164 however, there are still many challenges. The method proposed by the authors assumed that  
165 animals have a spatially-invariant movement, not taking into account other relevant factors,  
166 such as behavioural changes or habitat preferences of the species.

167 As distance sampling methods are based on collecting "real" absence data, cetacean  
168 distributional estimates are generally calculated using presence-absence algorithms. Some  
169 algorithms are widely used, such as Generalized Linear Models (GLMs) or Generalized  
170 Additive Models (GAMs) (e.g. Becker et al., 2014, Cañadas & Hammond, 2008, De Segura  
171 et al., 2008, Tepsich et al., 2014, Virgili et al., 2017). However, only recently, Fiedler et al.  
172 (2018) compared the use of algorithms using background data (Maxent) with the  
173 performance of GAMs to calculate distributional estimates of cetaceans in the Tropical  
174 Pacific Ocean area, finding non-significant differences between both approaches.

175 In the present study, we use 20 cetacean-based virtual species with distinctive habitat  
176 preferences and encounter rates to study how the use of different kinds of absences (or  
177 equivalent) affects modelling algorithms' performance. We aim to understand better how  
178 the presence of methodological absences in standardized survey datasets might influence



179 the accuracy of the predictions when using: (1) presence-absences, (2) presence-  
180 background and (3) presence-only modelling approaches. The study's goal is to describe  
181 potential method-biases and provide recommendations on how to account for them.

## 182 **METHODS**

### 183 **Study area**

184 The study area is located in the Azores archipelago, a group of North Atlantic oceanic is-  
185 lands located approximately 1,800 km west of Lisbon, Portugal. The region is strongly in-  
186 fluenced by the Gulf Stream and all its branches, creating a complex and highly dynamic  
187 structure. The system is affected by incoming meanders and filaments originating in the  
188 Gulf Stream and westward propagating eddies pinching off from the Azores Current  
189 (Caldeira & Reis 2017). Other factors, such as the archipelago's importance for retaining in-  
190 coming particles, might also play a role in local biological systems (Sala et al., 2015).

191

### 192 **Environmental variables**

193 A set of actual marine environmental variables was selected to represent the variability and  
194 dynamism of an oceanic system. Only three variables were chosen to facilitate the model-  
195 ling processes based on their reported influence on previous cetacean distribution studies in  
196 the area (e.g. Fernandez et al., 2019). These were divided into two thematic groups: static  
197 (little or no short-term variation—topographic variables) and dynamic (rapidly changing  
198 variables, such as sea surface temperature).

199

200 Two static variables were used: depth (derived from the digital elevation model of the  
201 EMODnet Bathymetry portal); and slope, calculated from the digital elevation model using

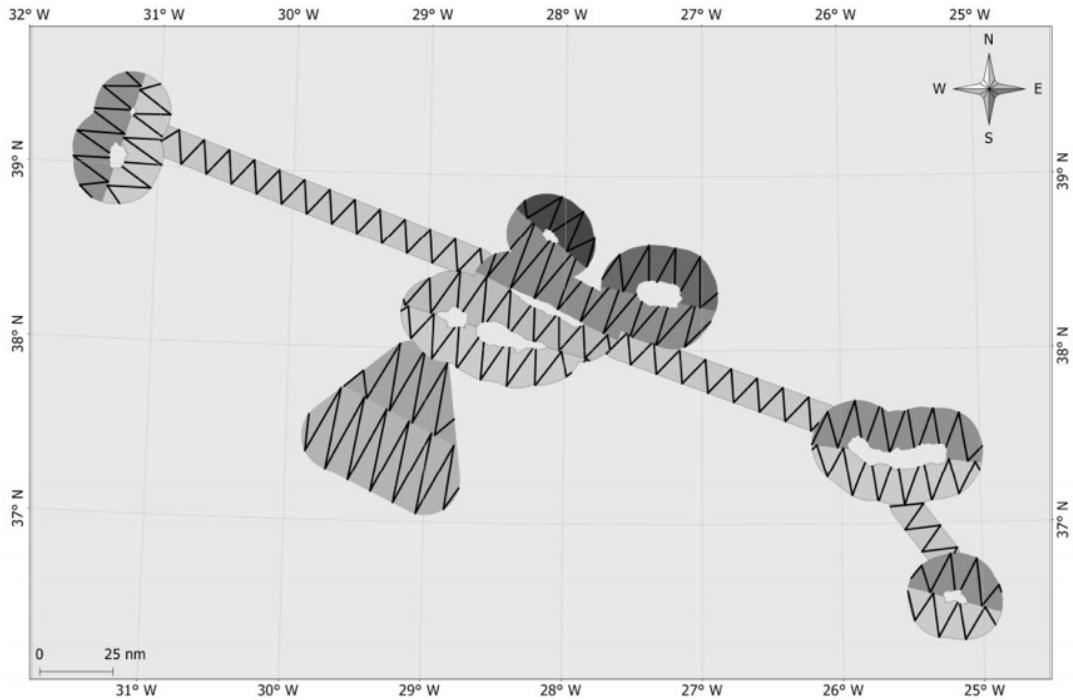
202 QGIS 3.2. All static variables were initially calculated at a spatial resolution of 0.5×0.5 km.  
203 Daily dynamic variables were derived from NASA's Multi-scale Ultra-high Resolution  
204 (MUR) Sea Surface Temperature (SST) dataset (downloaded on 01/12/2018 from [https://](https://mur.jpl.nasa.gov/)  
205 [mur.jpl.nasa.gov/](https://mur.jpl.nasa.gov/)), which merges many infrared and passive microwave datasets, gathered  
206 from satellites, into daily global maps at 1 km resolution corresponding to the sampling  
207 period (see survey design for further details). For the present study, we calculated 8-day es-  
208 timates using the means of SST daily values. Finally, all the layers were aggregated to 2×2  
209 km resolutions using a bilinear interpolation. We applied a variance inflation factor (VIF)  
210 approach as implemented in the R package usdm (Naimi, 2015) to test for collinearity, no  
211 issues were found.

212

### 213 **Survey design**

214 The virtual species' environmental and effort data mirrored the time frame of a simulated  
215 cetacean detection survey for the Azores archipelago. Surveys were performed around the  
216 islands and other vital areas, covering 39,316.9 km<sup>2</sup> per year. We used the transects de-  
217 signed by Faustino et al. (2010) (Fig 2) to last four months (12 weeks) per year, with four  
218 days of sampling per week over four years (2013-2016). Cruise speed was defined as 7.5  
219 knots during 8 hours each sampled day to avoid any area's resampling. Survey transects  
220 were mapped onto the 2 km grid that matched the environmental data.

221



222

223 **Figure 2.** Survey design with lines and complete study (Azores Archipelago) area used for  
 224 the models' projections and overlap metrics. Different colors refer to substrata to be sam-  
 225 pled (adapted from Faustino et al. 2010).

226

227 **Virtual species**

228

229 *Species definitions*

230 A total of 20 virtual species were created. Four different relationships with the environment  
 231 variables were used: (1) generalist species highly influenced by static variables; (2) special-  
 232 ist species highly influenced by static variables; (3) generalist species highly influenced by  
 233 dynamic variables; (4) specialist species highly influenced by dynamic variables. The spe-  
 234 cies were created using the Virtualspecies R package (Leroy et al., 2015) based on their (as-  
 235 sumed) relationship with daily environmental variables. For simplicity, bell-shaped func-

236 tions (Gaussian functions) and logistic functions were used to define the relations with the  
 237 environmental parameters (for full details of construction, see Appendix S1). Response  
 238 functions were combined using an additive approach with different weights for each vari-  
 239 able (Tab. 1), based on results found for other species in the same area by Fernandez et al.  
 240 (2018). Habitat suitability for each of the four general kinds of species (generalist static,  
 241 specialist static, generalist dynamic and specialist dynamic) was calculated every 8 days.

242

243 **Table 1.** Description of the different species with the information regarding their relation-  
 244 ship with the environmental variables and the encounter rates (encounters per 100km) used  
 245 for each scenario.

Species	Variable weights	Encounter rates (enc./ 100km)
Generalist static (group 1)	6Depth+3Slope+SST	1.4,1,0.8,0.6,0.2
Specialist static (group 2)	8Depth+Slope+SST	1.4,1,0.8,0.6,0.2
Generalist dynamic (group 3)	5SST+4Depth+Slope	1.4,1,0.8,0.6,0.2
Specialist dynamic (group 4)	6SST+2Depth+2Slope	1.4,1,0.8,0.6,0.2

246

247 Each day, environmental suitability was converted into presence/absence areas for all the  
 248 sampled regions, using a probabilistic approach suggested by Meynard and Kaplan (2013).  
 249 To perform this conversion, we calculated a logistic curve with  $\beta=0.6$  (representing the  
 250 curve's inflexion point) and  $\alpha=-0.05$  (which creates a logistic shape for the curve). Using  
 251 this approach, we forced the presences to be more likely in areas with suitability values  
 252 around or higher than 0.6 and allowed some individuals to occur in areas with low suitabil-  
 253 ity (sometimes even at values lower than 0.3). This represents a realistic scenario where

254 species can be observed moving through environmentally sub-optimal areas to access other  
255 suitable locations.

256 For each species, five different groups of sampling encounter rates were used per 100 km:  
257 1.4, 1, 0.6, 0.4 and 0.2, based on real-world data observations obtained by Silva et al.  
258 (2014) in the Azores. We assumed a perfect presence/absence detection scenario to sim-  
259 plify the analysis: all the groups present during the sampling were detected, although this is  
260 unusual for marine species (Katsanevakis et al., 2012). Assuming a value of 1 for the de-  
261 tectability, the encounter rate values could be linked with the species prevalence or how the  
262 species uses the suitable habitat. Even if the prevalence is a property of the data, it usually  
263 covaries with species ecology and range size, i.e., data for rare species usually show low  
264 prevalence scores, while widely-distributed species show the opposite (Jimenez-Valverde et  
265 al. 2009). Therefore, species with higher encounter rates will occupy more suitable grid  
266 cells (representing widely-distributed species) than those with lower encounter rates (rep-  
267 resenting rare species).

#### 268 *Sample occurrences*

269 Once presence/absence areas were delimited, we performed the last step of the virtual spe-  
270 cies generation process: sampling species occurrences. The occurrences points were calcu-  
271 lated based on the presence/absence areas previously delimited, each case's encounter rates,  
272 and the weighted environmental suitability values. Following Fernandez et al. (2017), we  
273 allowed the encounter rate to change through time related to the amount of suitable daily  
274 habitat per kilometre. This creates a more realistic scenario where, for dynamic species, en-  
275 counter rates will increase with appropriate conditions, simulating dynamic populations,  
276 where species are primarily present when conditions are good.

277

278 According to the niche modelling temporal resolution, daily presence points were pooled  
279 together with the effort and the consequent environmental variables (with 8-day means of  
280 SST) for each grid cell. A value of 1 was assigned for each sighting; when more than one  
281 sighting was found in a segment, only one presence point was used.

282

### 283 **Modelling techniques**

284

285 To understand how different types of absences affect the modelling processes, we per-  
286 formed the analysis using six different approaches, divided into three distinct groups related  
287 to the occurrences data sources' nature: presence-only, presence-background and presence-  
288 absence. Moreover, as characteristics of target species might also influence model perform-  
289 ance (Qiao et al., 2015), we used different algorithms to minimize this problem.

290

#### 291 *Presence-absence algorithms (PA)*

292 In this case, observed absences were treated as true absences. We used two kinds of pres-  
293 ence-absence algorithms: Generalized Linear Models (GLMs) and Generalized Additive  
294 Models (GAMs). Both methods are based on the correlation of presences and genuine the-  
295 oretical absences (or pseudo-absences) with environmental variables. Therefore, these mod-  
296 els assume that environmental conditions were not suitable for survival if the species were  
297 not found. We fitted binomial GLMs with quadratic terms using a logit link using R (ver-  
298 sion 3.4.0; R Core Team, 2017). Two different approaches were used for the GAMs fitting  
299 process using the "mgcv" package (Wood, 2011): (1) using all presences and all theoretical  
300 absences pooled together, and (2) running ten iterations with all presences and 1000 theor-  
301 etical absences selected randomly on each run, mimicking the “absences zones” approach

302 used by Derville et al. 2016. We used a logit link for both approaches and allowed a max-  
303 imum of three degrees of freedom for each spline to limit over-fitting (Becker et al., 2014).  
304 Furthermore, the model's effort was included as an offset term to account for variations in  
305 segment length.

306

#### 307 *Presence-background algorithms (PB)*

308 For this scenario, observed absences were treated as a sample of the conditions available in  
309 the region. Maxent uses presence points and background samples pooled together with en-  
310 vironmental variables to calculate the potential habitat suitability (Phillips et al., 2017).  
311 Background points represent environmental conditions available for the species (Phillips et  
312 al., 2009). Background includes randomly selected pixels where the species have been ob-  
313 served and pixels where the species was not observed (Phillips et al., 2009). Maxent (ver-  
314 sion 3.4.1) was implemented using the R package "dismo" (Hijmans et al., 2017). We fitted  
315 two different Maxent approaches: (1) a classical approach with automatic features and (2) a  
316 more simplistic approach with only linear and quadratic features. We decided to test the  
317 second approach due to the virtual species simplistic response curves designed for the  
318 present study. We applied a regularization factor of 2 to avoid over-fitting (Fiedler et al.,  
319 2018). Occurrences and background data were pooled into a table in the sample-with-data  
320 format to input Maxent in a Sightings With Data (SWD) format.

321

#### 322 *Presence-only algorithms (PO)*

323 Presence-only algorithms are solely based on the observed occurrences, not considering the  
324 absences registered. For the present study, we used the BIOCLIM (Busby, 1986, Booth et

325 al., 2014) and DOMAIN (Carpenter et al., 1993) implementation on the R package  
326 "dismo".

327

### 328 **Validation techniques**

329 To assess the accuracy of the models produced, we used a variety of validation techniques:  
330 one index based on the Area Under the Curve (AUC), the response curves plots from the  
331 different modelling algorithms (a plot of species presence in relation to a changing environ-  
332 mental variable) and a group of indices based on the comparison of models predictions with  
333 the suitability of the virtual species, in geographic and environmental space. Ideally, dis-  
334 crimination measures such as AUC should be obtained with the use of an independent data-  
335 set (Peterson et al., 2011) or at least with the use of spatial (or temporal) bins to perform a  
336 binned cross-validation (Radosavljevic & Anderson 2014). However, this is still not com-  
337 mon practice, and there are many examples of cetacean studies using AUC based on ran-  
338 domized testing fold selection (e.g. Fiedler et al., 2018, Virgili et al., 2017). Therefore, to  
339 understand how these metrics explain the model accuracy, we calculated the test AUC with  
340 a random 5-fold approach (approx. 25% of dataset for testing). These metrics were calcu-  
341 lated using the R "dismo" package.

342

343 The other two validation techniques were based on the niche overlap between virtual and  
344 modelled virtual species, both in geographic and environmental space. These analyses  
345 quantify the accuracy of the model predictions in comparison with "real" virtual species  
346 niches. Firstly, we calculated the Schoener's D (Schoener, 1968) in the geographical space  
347 (G-space) using the R package ENMtools (Warren et al., 2017). Schoener's D ranges from  
348 0 (dissimilarity) to 1 (similarity) (Warren et al., 2017). The comparison of the model's geo-



349 graphical projections measures the predictions' accuracy using the observed environmental  
350 conditions. This analysis was performed using the comparison of the models' projection  
351 with the actual theoretical suitability for each species in all the study area (including many  
352 places outside the surveyed regions but with similar environmental conditions). A fuzzy  
353 similarity index was applied to the G-space to find differences between projection out-  
354 comes. These indices consider the fuzziness of the location, going beyond site-by-site com-  
355 parison and giving partial credit to neighbouring sites (Barbosa, 2015). We used a fuzzy  
356 similarity index based on Baroni-Urbani & Buser's (1976) index, which accounts for both  
357 shared presences and shared absences but gives greater weight to presences. Analysis was  
358 implemented using the fuzzySim R package (Barbosa, 2015).

359

360 Working in the geographical dimension can be problematic because the measured niche  
361 overlap is likely to vary depending on the extent and distribution of environmental gradi-  
362 ents in the study area and unquantified statistical artefacts related to model fitting (Broenni-  
363 mann et al., 2012). To overcome the biases associated with the geographical dimension,  
364 Broennimann et al. (2012) proposed to perform niche overlap tests directly in environ-  
365 mental space, allowing the use of all potential combinations of variables in an n-dimen-  
366 sional space. This process was applied using the ENMTools (Warren et al., 2021), drawing  
367 Latin hypercube samples from the space of all possible combinations of environmental  
368 variables given the minimum and maximum of each variable within the training region  
369 (Warren et al., 2019). The Schoener's D index and the Spearman correlation rank were ap-  
370 plied to test for differences between modelling scenarios.

371

372

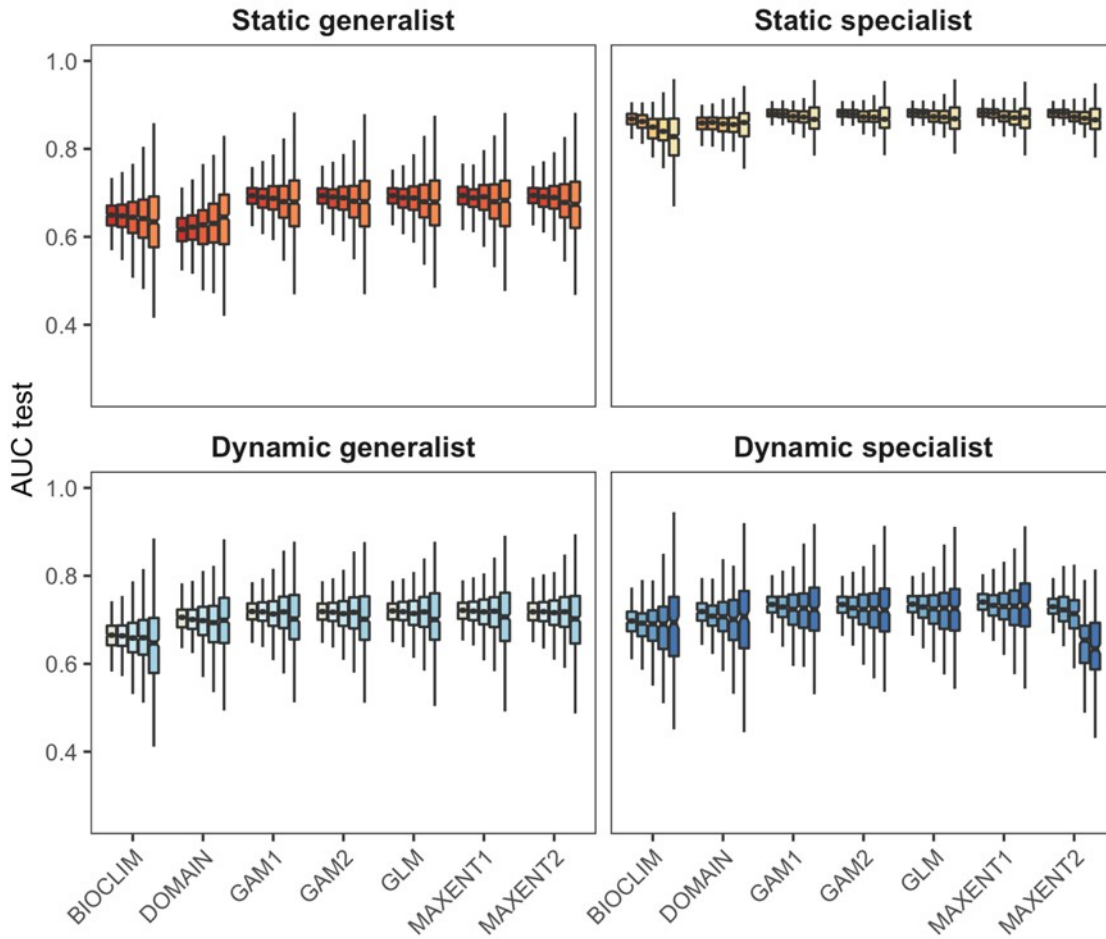
373

374

## 375 **RESULTS**

376

377 A total of 14000 models were constructed corresponding to the 20 virtual species, 7 algo-  
378 rithms types, and 100 iterations for each scenario. We found differences in the results ac-  
379 cording to the species analysed, the encounter rates and the validation metric used. In con-  
380 trast, similar results between algorithms were obtained when using the AUC test metric  
381 (Fig. 3). Nevertheless, higher variability in the accuracy was observed in the results with  
382 lower encounter rates. The same effect is noticeable when using other validation tech-  
383 niques, especially for the lowest encounter-rate scenarios (0.2 encounters per 100km).  
384 Moreover, AUC test values for specialists are higher than for generalist species, particularly  
385 in static species.



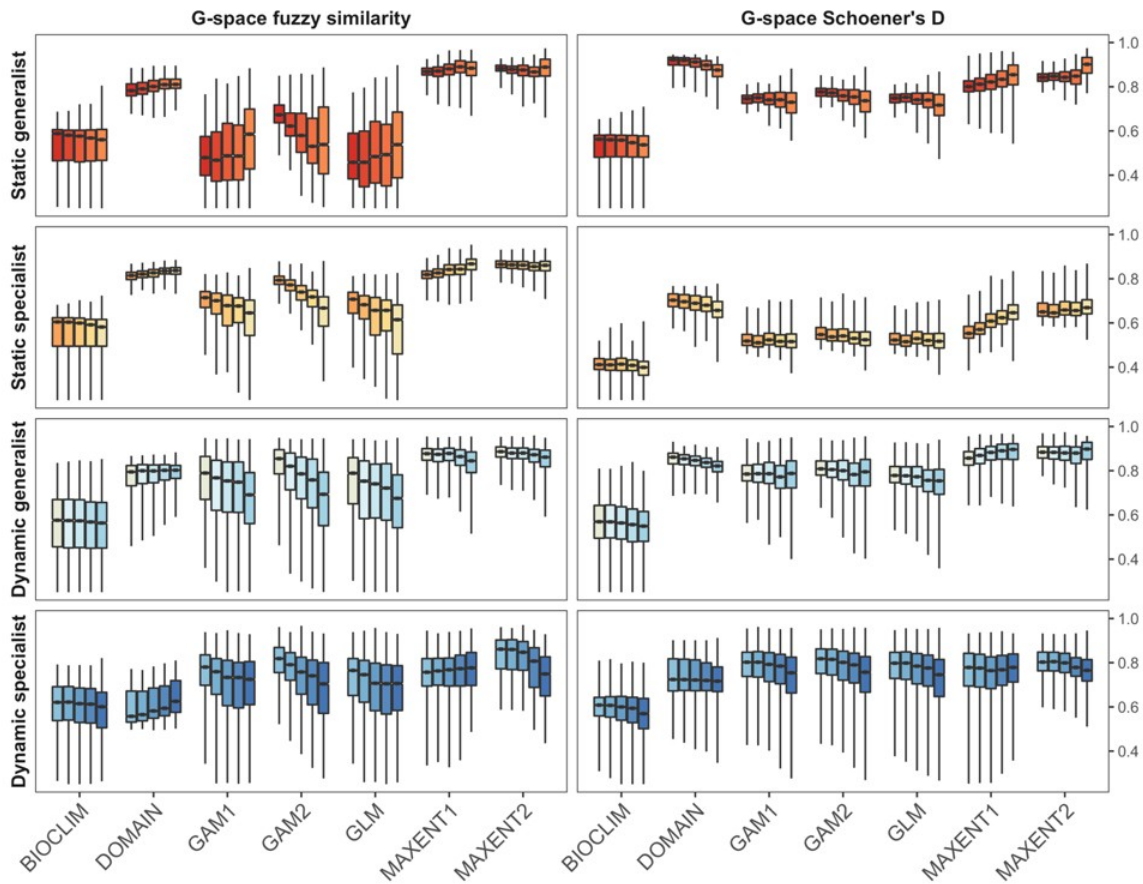
386

387 **Figure 3.** Area Under the Curve (AUC) using testing dataset values for 4 different species  
 388 (generalistic static, specialist static, generalist dynamic, specialist dynamic) with different  
 389 encounters rate (1.4, 1, 0.8, 0.6 and 0.2 encounters per 100km sampled; values go from left  
 390 to right). Mean and standard deviation values are presented for each of the six modelling  
 391 approaches tested (BIOCLIM, DOMAIN, GLM, GAM1, GAM2, Maxent1 and Maxent2).

392

393 The two tests performed on the geographical space highlighted differences between model-  
 394 ling techniques (Fig 4). Schoener's D index presented lower overlap values for the special-  
 395 ist species than the generalists. However, this pattern is not present when using the Fuzzy

396 similarity test. For the different algorithms, the lowest overlap values were consistently ob-  
 397 tained for BIOCLIM, while the use of DOMAIN resulted in high overlap values for both  
 398 indexes (Fig 4). Together with DOMAIN, the MAXENT algorithm predictions resulted in  
 399 higher overlap and similarity indexes than presence-absence algorithms (GAM2, GAM2  
 400 and GLM). Nevertheless, this trend is not so evident in the Dynamic specialist species,  
 401 where overlap indexes for all models in G-space are highly variable.



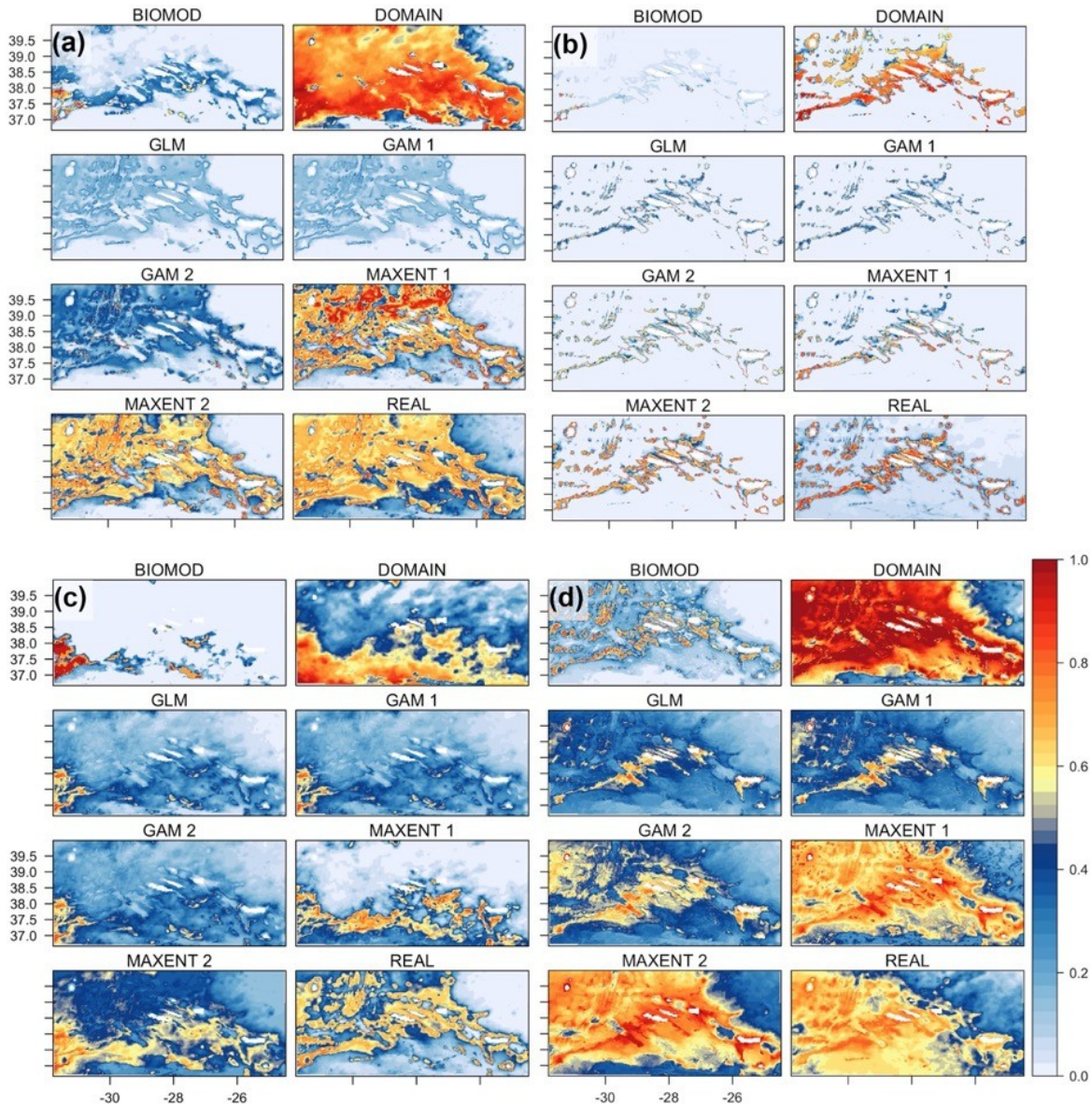
402  
 403 **Figure 4.** Fuzzy similarity and Schoener's D overlap index values on the geographic space  
 404 between the model projections and the theoretical "real" suitability maps. Results are  
 405 grouped by species type (static generalist, static specialist, dynamic generalist, dynamic  
 406 specialist). Each bar indicates the results for the different encounter rates tested (1.4, 1, 0.8,

407 0.6 and 0.2 encounters per 100km sampled; values go from left to right) for each kind of  
408 modelling algorithm specified on the X-axis.

409

410 The projected maps for the 4 selected species (Fig 5) depict the poorer performance of the  
411 predictions made using the BIOCLIM algorithm. On the other hand, the DOMAIN algo-  
412 rithm over-predicts some areas. The PA models (GLM, GAM1 and GAM2) tend to under-  
413 predict suitable areas while MAXENT produced the best results (but still with some areas  
414 over-predicted). All the algorithms showed difficulties predicting the low suitability ranges  
415 gradient, which is especially clear when looking at the lower suitability values (blue gradi-  
416 ent) for static species (Fig 5a and 5b).

417



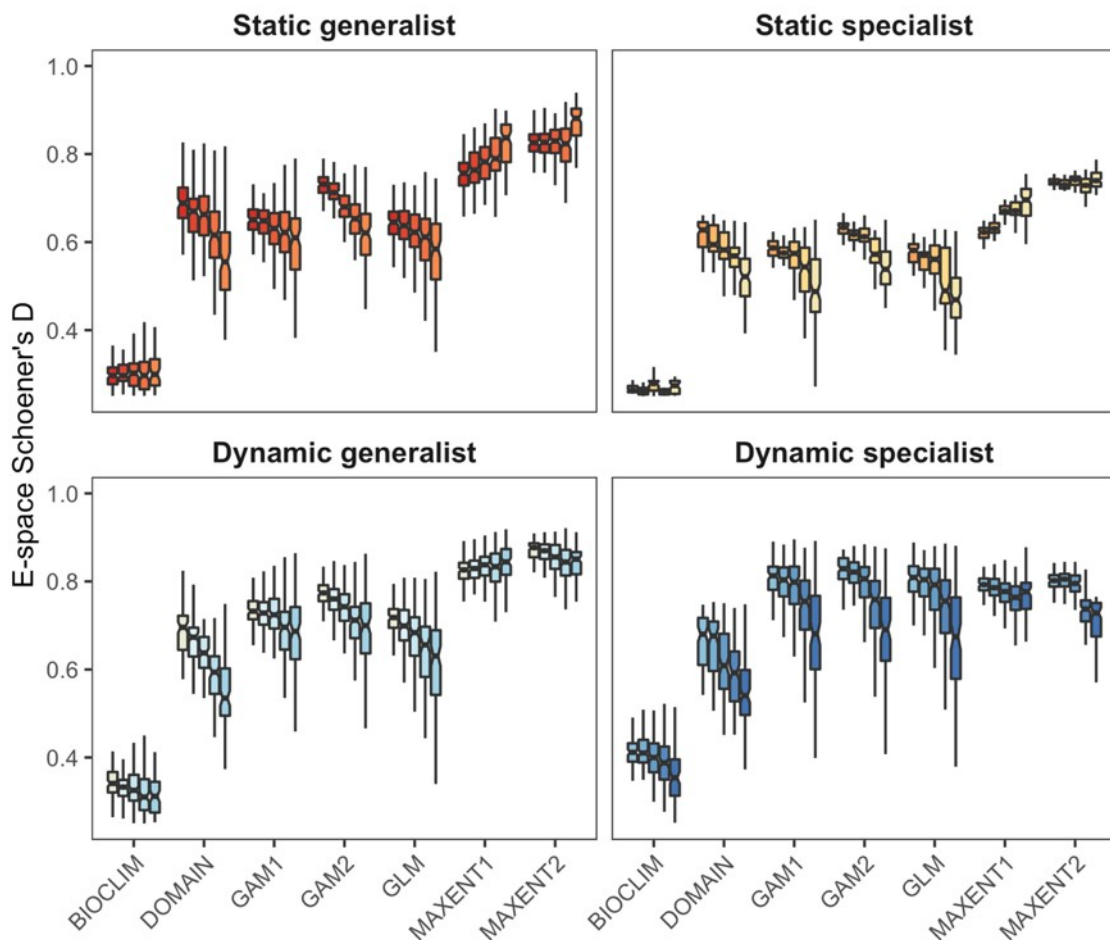
418

419

420 **Figure 5.** Predicted suitability values for four different species (the species with an en-  
 421 counter rate of 1.4 encounters per 100km sampled for each type) on a randomly selected  
 422 week, comparing all the modelling techniques used. (a) (top left) species 1, a static general-  
 423 ist species; (b) (top right) species 6, a static specialist species; (c) (bottom left) species 11, a  
 424 dynamic generalist species; (d) (bottom right) species 16, a dynamic specialist species.  
 425 Suitability ranges from 0 (white) to 1 (red).



426 In the environmental space (E-space) evaluation, models constructed using MAXENT pro-  
 427 duced higher Schoener's D overlap indexes (Fig. 6). Only the Dynamic specialist species  
 428 values seem similar among all algorithms tested; nevertheless, MAXENT models had much  
 429 lower standard deviation values. DOMAIN predictions were equal or worse than those of  
 430 PA algorithms. Species with lower encounter rates produced the worst results, resulting in  
 431 higher standard deviation values.  
 432



433  
 434 **Figure 6.** Schoener's D overlap index values for the environmental space between the  
 435 model outputs and the theoretical "real" species relation with environmental variables. Re-

436 sults are grouped by species type (static generalist, static specialist, dynamic generalist, dy-  
437 namic specialist). Each bar indicates the results for the different encounter rates tested (1.4,  
438 1, 0.8, 0.6 and 0.2 encounters per 100km sampled; values go from left to right) for each  
439 kind of modelling algorithm specified on the X-axis.

440

441 The response curves (Fig. S5-S24) reflect a similar pattern to those observed in the other  
442 validation analysis. When considering the weighting of the variables for the different  
443 species, the MAXENT approaches performed better in general, with lower variability in the  
444 results and curves more similar to those expected. Nevertheless, for the specialist species  
445 (Figs.S10-S14 and Figs.S20-S24) the curves for the regression methods (GLM and GAMs)  
446 also showed a good performance. Curves for DOMAIN and BIOCLIM were, in general,  
447 overestimated and unable to detect the different weighting of the variables.

448

449 Finally, the linear approach for MAXENT (MAXENT2 in the figures) for the static special-  
450 ists' species produced better estimates. In contrast, the classic MAXENT approaches varied,  
451 producing similar results to the PA algorithms when working with high encounter rates. PB  
452 approaches and PA produced similar results for the dynamic specialists' species when using  
453 high encounter rates. However, with low encounter rates, the standard deviation values for  
454 the PB were lower than all the other approaches.

455

456 The use of different settings for GAMs (GAM1 and GAM2) and MAXENT (MAXENT1  
457 and MAXENT2) proved to be effective in some cases, with different results observed de-  
458 pending on the settings used, as can be observed in all the results obtained (Fig. 4, 5 and 6).

459



## 460 **DISCUSSION**

461

462 Overall, depending on how observed absences were treated, analyses showed apparent dif-  
463 ferences when using distinct algorithms to characterize the ecological niche for the simu-  
464 lated species. The different validation techniques allowed us to obtain an accurate idea of  
465 the methods' weaknesses and strengths. For example, the two metrics calculated in G-space  
466 (Fig. 4), a more classical approach (Schoener's D index) and a more complex approach  
467 (Fuzzy similarity), permitted us to understand the models' projections differences better.  
468 Using E-space for validation proved beneficial as it reinforced some subtle tendencies  
469 barely visible with the G-space methods.

470

471 The overlap tests (both in environmental and geographical space) indicate that the model-  
472 ling approaches treating observed absences as background data could equally (or better) de-  
473 termine the species niche than when considering the non-observation as a real absence (PA  
474 methods) or when considering no absences at all (PO approach). This might be related to  
475 specific ecological traits; highly mobile species, such as cetaceans or other pelagic marine  
476 megafauna, might use the habitat versatily. Oceans are in constant flux; the physical forc-  
477 ing of the oceanographic seascape spans multiple spatial and temporal scales. Oceano-  
478 graphic features that both predators and preys follow, like fronts and eddies, move dynami-  
479 cally across similar scales (Steele, 1991). Therefore, even the so-called specialists' large  
480 marine predators might have flexible and dynamic habitat use when looking for prey  
481 patches. Moreover, animal migrations and foraging or breeding aggregations track dynamic  
482 oceanographic features across multiple scales (Block et al. 2011). Consequently, the non-  
483 observation of a species in a specific location and time might easily constitute a methodolo-

484 gical absence. As Lobo et al. (2010) suggested, this can lead to obtaining absences in suit-  
485 able environmental locations, especially in areas close to recorded presences. Therefore, a  
486 very intense sampling schema (unrealistic) would be required to obtain a genuine unbiased  
487 absence when working with these species.

488

489 Ideally, when trying to unveil or predict the occupied niche (or the distribution) of a spe-  
490 cies, actual absence data and PA models might be the best theoretical option (Sillero,  
491 2011). As we discussed, our results suggest that PA models should not be applied when  
492 true absences are not available. PA models attempt to calculate accurate distribution estim-  
493 ates (distinguishing between occupied and non-occupied habitats), producing, in some  
494 cases, unrealistic and over-fitted predictions. While other methods, such as presence-back-  
495 ground methods, are more flexible (Peterson et al., 2011). It is precisely this flexibility that  
496 might be a better choice when working with highly motile species, as when working in dy-  
497 namic environments and mobile species, the non-observation of a species in a given loca-  
498 tion and time should not be considered as an absence but instead considered as a sample of  
499 the whole study area; which agrees with the definition of a background sample (Phillips et  
500 al., 2009). Summing the G and E space results resulted in a better (or at least equal) per-  
501 formance of PB approaches over all the other methods tested. Interestingly, Fiedler et al.  
502 (2018) found that Maxent can produce models similar to GAM presence-absence models  
503 when background data points are selected from observed absences. It is important to notice  
504 that the cited study only used AUC metrics and a point biserial correlation to compare the  
505 outputs. Instead, we decided to use a broader set of metrics (in the G and E space) to com-  
506 pare the model predictions, which allowed us to find some differences between modelling  
507 approaches that might be missed when using more simplistic comparison methods.

508

509 Interpretations based solely on geographic space hold relatively limited information and  
510 should be taken with caution; models should instead be analysed in environmental and geo-  
511 graphic spaces (Escobar et al., 2018). Comparing predictions in environmental space (Fig.  
512 6) allowed us to overcome some of the limitations derived from the structure and composi-  
513 tion of environmental variables in the specific region of study highlighting differences  
514 between modelling approaches that would otherwise remain invisible (Warren et al., 2019).  
515 Summing the G and E space results resulted in a better (or at least equal) performance of  
516 PB approaches over all the other methods tested.

517

518 An interesting pattern observable in E-space is that, for generalist species, treating the ob-  
519 served absences as background data seemed to perform consistently better (with higher  
520 overlap values and lower standard deviations), agreeing with Evangelista et al. (2018) find-  
521 ings. However, different patterns emerge for specialist species, with different results de-  
522 pending on the species, encounter rate, and modelling algorithm used.

523

524 Some of the differences observed between modelling techniques might be related to the  
525 niche breadth of the target species. Broad realised niches (generalist species; species 1 and  
526 3) are more challenging to model than narrow niches (specialist species; species 2 and 4),  
527 with marked differences in performance when using different algorithms (Qiao et al., 2015,  
528 Connor et al., 2018). The present study found that the differences between algorithms are  
529 less visible for specialist species, especially for static ones (species 4). When the available  
530 suitable area is more restricted and stable through time, the sampling effort needed to ob-  
531 tain genuine environmental or contingent absences is lower. For example, suppose a ceta-

532 cean species occurs only in coastal areas, and we sample coastal and non-coastal areas. In  
533 that case, the chance of obtaining true absences is higher than if the species occur over a  
534 broader range of conditions. Therefore, using observed absences as real absences and ap-  
535 plying presence-absence models might be a good choice in this scenario. Nevertheless, our  
536 results showed that even if PA models performed well, PB methods (and even PO in some  
537 cases) performed equally well.

538

539 Furthermore, our results agree with other findings on the importance of model settings and  
540 calibration (Warren et al., 2014). We obtained different results when using the same type of  
541 algorithm but with different settings or parameter combinations. Parameter choice can pro-  
542 duce dramatic influences on model output. We found that using the GAM2 approach (using  
543 the sampling approach from “absences zones”) produced better results than a classical  
544 GAM approach. The same was observed with the two settings used for MAXENT models.  
545 Hence, besides testing different algorithms, it is also critical to build models with different  
546 settings, using tools to perform detailed development of ecological niche models, such as  
547 kuenm (Cobos et al., 2019) or ENMeval (Kass et al., 2021).

548

## 549 **Conclusion**

550

551 As a result of the experiences and analysis performed here, we cannot recommend one  
552 unique way of considering observed absences for niche modelling when working with  
553 highly mobile species. Our results suggest that in highly mobile species, an excellent pre-  
554 cautionary approach would be to consider the non-observation of a species as a background  
555 point (a sample of the conditions available in the sampled area) rather than an absence.

556

557 However, as Quiao et al. (2015) pointed out, there are no silver bullets, and therefore there  
558 is no fixed recipe that will work for all the cases. In some cases (such as working with static  
559 species with high prevalence), the absences encountered during the sampling process might  
560 be a good approximation to real absences. Therefore, PA methods would be an optimal  
561 choice. However, sufficient ecological knowledge of the species may be unavailable in  
562 many cases, especially when working with marine megafauna. In such cases, using a PB  
563 method might be a good starting point, which ideally should be complemented with PA and  
564 PO models. Afterwards, evaluation tests should be performed, comparing the outputs from  
565 the different algorithms tested in G and E space, as suggested by previous studies (Escobar  
566 et al., 2018, Warren et al., 2019).

567

568 Ideally, we would like to obtain the probability of finding a species on a specific site when  
569 modelling distributions for management purposes. In the present study, we focused on tran-  
570 sect survey data; however other methods, such as tracking, could help obtain better insights  
571 into the species movements and habitat preferences. Scales et al. (2016) used movement  
572 simulations of blue whales to better investigate the potential of tracking data for predictive  
573 habitat modelling. Fine-scale movement data might provide an excellent opportunity to ob-  
574 tain distributional models at finer temporal resolutions and, therefore, improve management  
575 applications, such as collision risk assessments (Blondin et al. 2020). Moreover, PA mod-  
576 els' estimates are based on occurrence and absence data collected on a fixed time frame,  
577 and therefore they predict the probability of occurrence. Areas with high suitability values  
578 do not guarantee that the species will always be present. With highly motile species, occu-  
579 pancy rates might vary even more. They have broader home ranges, and these movements

580 can be related to many factors, such as competition, food availability or behaviour spe-  
581 cificities. However, including these covariates is highly complex and requires continuous  
582 recording systems and further study. Long-term moored acoustic systems might help an-  
583 swer some of these questions, improving distributional and occupancy estimates (e.g.  
584 Fraiser et al., 2021).

585

## 586 **Acknowledgements**

587

588 We thank Claudia Faustino for the shapefiles to simulate the sampling transects used in the  
589 manuscript. Cetacean illustrations used in Figure 1 were made by E. Berninsone. This study  
590 was partially supported by: (i) INTERTAGUA, MAC2/1.1.a/385 funded by MAC INTER-  
591 REG 2014-2020, (ii) Oceanic Observatory of Madeira throughout the project M1420-01-  
592 0145-FEDER-000001-OOM, and (iii) Fundação para a Ciência e Tecnologia (FCT), Por-  
593 tugal, through the strategic project UID/MAR/04292/2020 granted to MARE UI&I. NS is  
594 supported by a CEEC2017 contract (CEECIND/02213/2017). Authors at ZSL (CY) and  
595 their research are supported by funding from Research England.

596

597 *MF conceived the ideas and designed methodology; MF, NS and CY analyzed the data;*  
598 *MF, NS and CY led the writing of the manuscript. All authors contributed critically to the*  
599 *drafts and gave final approval for publication.*

600

## 601 **REFERENCES**

602

603 Barbosa, A.M. (2015), fuzzySim: applying fuzzy logic to binary similarity indices in eco-  
604 logy. *Methods Ecol Evol*, 6, 853-858. doi:[10.1111/2041-210X.12372](https://doi.org/10.1111/2041-210X.12372)

605

606 Baroni U.C., Ruffo, S. & Taglianti, V.A. (1978). Materiali per una biogeografia italiana  
607 fondata su alcuni generi di coleotteri Cicindelidi, Carabidi e Crisomelidi. *Memorie della*  
608 *Società Entomologica Italiana*, 56, 35– 92.

609

610 Block, B.A., Jonsen, I.D., Jorgensen, S.J., Winship, A.J., Shaffer, S.A., Bograd, S.J. et al.  
611 (2011) Tracking apex marine predator movements in a dynamic ocean. *Nature*, 475, 86–90.

612

613 Blondin, H., Abrahms, B., Crowder, L. B., & Hazen, E. L. (2020). Combining high tempo-  
614 ral resolution whale distribution and vessel tracking data improves estimates of ship strike  
615 risk. *Biological Conservation*, 250, 108757.

616

617 Becker, E.A., Forney, K.A., Foley, D.G., Smith, R.C., Moore, T. J., & Barlow, J. (2014).  
618 Predicting seasonal density patterns of California cetaceans based on habitat models. *En-*  
619 *dangered Species Research*, 23(1), 1-22.

620

621 Booth T.H., Nix H.A., Busby J.R., & Hutchinson M.F. (2014) BIOCLIM: the first species  
622 distribution modelling package, its early applications and relevance to most current MaxEnt  
623 studies. *Divers Distrib*, 20, 1–9.

624

625 Boria, R.A., Olson, L.E., Goodman, S.M., & Anderson, R.P. (2014). Spatial filtering to re-  
626 duce sampling bias can improve the performance of ecological niche models. *Ecological*  
627 *modelling*, 275, 73-77.

628

629 Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz,  
630 N.G., Thuiller, W., Fortin, M.J., Randin, C., Zimmermann, N.E., Graham, C.H. & Guisan,  
631 A. (2012), Measuring ecological niche overlap from occurrence and spatial environmental  
632 data. *Global Ecology and Biogeography*, 21, 481-497. doi:[10.1111/j.1466-](https://doi.org/10.1111/j.1466-8238.2011.00698.x)  
633 [8238.2011.00698.x](https://doi.org/10.1111/j.1466-8238.2011.00698.x)

634

635 Brotons, L., Thuiller, W., Araújo, M.B., & Hirzel, A.H. (2004). Presence-absence versus  
636 presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4),  
637 437-448.

638

639 Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., & Thomas, L.  
640 (2001). Introduction to distance sampling: estimating abundance of biological populations.  
641 Oxford University Press, Oxford.

642

643 Busby, J.R. (1986) A biogeographical analysis of *Nothofagus cunninghamii* (Hook.) Oerst.  
644 in southeastern Australia. *Aust. J. Ecol.* 11, 1-7.

645

646 Caldeira, R., & Reis, J.C. (2017). The Azores confluence zone. *Frontiers in Marine Sci-*  
647 *ence*, 4, 37.

648



649 Cañadas, A., & Hammond, P.S. (2008). Abundance and habitat preferences of the short-  
650 beaked common dolphin *Delphinus delphis* in the southwestern Mediterranean: implica-  
651 tions for conservation. *Endangered species research*, 4(3), 309-331.

652

653 Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN — A flexible modeling procedure  
654 for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2,  
655 667–680.

656

657 Cobos M.E., Peterson A.T., Barve N. & Osorio-Olvera L. (2019). kuenm: an R package for  
658 detailed development of ecological niche models using Maxent. *PeerJ* 7, e6281

659

660 Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J., Wang, F., & Liu, J.  
661 (2018). Effects of grain size and niche breadth on species distribution modeling. *Eco-*  
662 *graphy*, 41(8), 1270-1282.

663

664 Dennis, R.L., Sparks, T.H., & Hardy, P.B. (1999). Bias in butterfly distribution maps: the  
665 effects of sampling effort. *Journal of Insect Conservation*, 3(1), 33-42.

666

667 Derville, S., Torres, L. G., Iovan, C., & Garrigue, C. (2018). Finding the right fit: Compara-  
668 tive cetacean distribution models using multiple data sources and statistical approaches. *Di-*  
669 *versity and Distributions*, 24(11), 1657-1673.

670

671 De Segura, A.G., Hammond, P.S., & Raga, J.A. (2008). Influence of environmental factors  
672 on small cetacean distribution in the Spanish Mediterranean. *Journal of the Marine Biolo-*  
673 *gical Association of the United Kingdom*, 88(6), 1185-1192.

674

675 Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J.,  
676 Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Man-  
677 ion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, M.J., Peterson, A.T., Phillips,  
678 S.J., Richardson, K., Scachetti-Pereira, R.E., Schapire, R., Soberón, J., Williams, S., Wisz,  
679 M.S., & Zimmermann, N.E. (2006). Novel methods improve prediction of species' distri-  
680 butions from occurrence data. *Ecography*, 29(2), 129-151.

681

682 Escobar, L.E., Qiao, H., Cabello, J., & Peterson, A.T. (2018). Ecological niche modeling  
683 re-examined: A case study with the Darwin's fox. *Ecology and evolution*, 8(10), 4757-4770.

684

685 Evangelista, P. H., Kumar, S., Stohlgren, T. J., Jarnevich, C. S., Crall, A. W., Norman III, J.  
686 B., & Barnett, D. T. (2008). Modelling invasion for a habitat generalist and a specialist  
687 plant species. *Diversity and Distributions*, 14(5), 808-817.

688

689 Faustino, C.E.S., Silva, M.A., Marques, T.A. & Thomas, L. (2010). Designing a shipboard  
690 line transect survey to estimate cetacean abundance off the Azores archipelago. *Ar-*  
691 *quipélago - Life and Marine Sciences*, 27, 49-58.

692

693 Fernandez, M., Yesson, C., Gannier, A., Miller, P. I., & Azevedo, J. M. (2017). The im-  
694 portance of temporal resolution for niche modelling in dynamic marine environments.  
695 *Journal of Biogeography*, 44(12), 2816-2827.

696

697 Fernandez, M., Yesson, C., Gannier, A., Miller, P. I., & Azevedo, J. M. N. (2018). A matter  
698 of timing: how temporal scale selection influences cetacean ecological niche modelling.  
699 *Marine Ecology Progress Series*, 595, 217-231.

700

701 Fiedler P.C., Redfern J.V., Forney K.A., Palacios D.M., Sheredy C., Rasmussen K., García-  
702 Godos I., Santillán L., Tetley M.J., Félix F., & Ballance L.T. (2018). Prediction of large  
703 whale distributions: a comparison of presence–absence and presence-only modeling tech-  
704 niques. *Frontiers in Marine Science*, 5, 419.

705

706 Franklin J (2010) Mapping Species Distributions. Cambridge University Press

707

708 Frasier, K.E., Garrison, L.P., Soldevilla, M.S., Wiggins, S.M., & Hildebrand, J.A. (2021).  
709 Cetacean distribution models based on visual and passive acoustic data. *Scientific reports*,  
710 11(1), 1-16.

711

712 Glennie, R., Buckland, S.T., Langrock, R., Gerrodette, T., Ballance, L.T., Chivers, S.J., &  
713 Scott, M. D. (2020). Incorporating animal movement into distance sampling. *Journal of the*  
714 *American Statistical Association*, 1-17.

715

716 Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A.T. (2004). New devel-  
717 opments in museum-based informatics and applications in biodiversity analysis. *Trends in*  
718 *ecology & evolution*, 19(9), 497-503.

719

720 Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sam-  
721 ple size and species characteristics on performance of different species distribution model-  
722 ing methods. *Ecography*, 29(5), 773-785.

723

724 Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., & Hijmans, M.R. J. (2017). Package  
725 ‘dismo’. *Circles*, 9(1), 1-68.

726 Hortal, J. (2008). Uncertainty and the measurement of terrestrial biodiversity gradients. *J.*  
727 *Biogeogr.* 35, 1355-1356.

728 Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008). Not as Good as They Seem: The  
729 Importance of Concepts in Species Distribution Modelling. *Divers. Distrib.*, 6 (14), 885–  
730 890.

731 Jiménez-Valverde, A., J. M. Lobo, & J. Hortal. (2009). The effect of prevalence and its  
732 interaction with sample size on the reliability of species distribution models. *Comm. Ecol.*  
733 10.2:196-205.

734 Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A.,  
735 Soley-Guardia, M., & Anderson, R.P. (2021). ENMeval 2.0: Redesigned for customizable  
736 and reproducible modeling of species’ niches and distributions. *Methods in Ecology and*  
737 *Evolution*. <https://doi.org/10.1111/2041-210X.13628>

738 Katsanevakis, S., Weber A., Pipitone C., Leopold M., Cronin, M., Scheidat, M., Doyle,  
739 T.K., Buhl-Mortensen, L., Buhl-Mortensen, P., D'Anna, G., de Boois, I., Dalpadado, P.,  
740 Damalas, D., Fiorentino, F., Garofalo, G., Giacalone, V.M., Hawley, K.L., Issaris, Y.,  
741 Jansen, J., Knight, C.M., Knittweis, L., Kröncke, I., Mirto, S., Muxika, I., Reiss, H.,  
742 Skjoldal, H.R. & Vöge S. (2012). Monitoring marine populations and communities: meth-  
743 ods dealing with imperfect detectability. *Aquatic Biology*, 16(1), 31-52.

744 Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder,  
745 V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M. & Cheyne, S.M. (2013). The  
746 importance of correcting for sampling bias in MaxEnt species distribution models.  
747 *Diversity and Distributions*, 19(11), 1366-1379.

748 Leatherwood, S., & Ljungblad, D.K. (1979). Nighttime swimming and diving behavior of a  
749 radio-tagged spotted dolphin, *Stenella attenuata*. *Cetology*, 34, 1-6.

750

751 Leroy, B., Meynard, C.N., Bellard, C., & Courchamp, F. (2016). Virtualspecies, an R pack-  
752 age to generate virtual species distributions. *Ecography*, 39(6), 599-607.

753

754 Lobo, J.M., Baselga, A., Hortal, J., Jiménez-Valverde, A., & Gómez, J.F. (2007). How does  
755 the knowledge about the spatial distribution of Iberian dung beetle species accumulate over  
756 time? *Diversity and Distributions*, 13(6), 772-780.

757

758 Lobo, J.M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences  
759 and their importance in species distribution modelling. *Ecography*, 33(1), 103-114.

760

761 Meynard, C.N., & Kaplan, D. (2013). Using virtual species to study species distributions  
762 and model performance. *Journal of Biogeography*, 40(1), 1-8.  
763

764 Naimi, B. (2015). usdm: Uncertainty Analysis for Species Distribution Models. R Package  
765 Version 1.  
766

767 Perrin, W.F., Evans, W.E., & Holts, D.B. (1979). Movements of pelagic dolphins (*Stenella*  
768 spp.) in the eastern tropical Pacific as indicated by results of tagging, with summary of tag-  
769 ging operations, 1969–76. NOAA Tech. Rep. NMFS No. SSRF-737.  
770

771 Peterson, A.T., Soberón, J., Pearson, R. G., Anderson, R.P., Martínez-Meyer, E., Na-  
772 kamura, M., & Araújo, M.B. (2011). Ecological niches and geographic distributions. –  
773 Princeton Univ. Press.  
774

775 Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., & Blair, M.E. (2017). Opening  
776 the black box: An open-source release of Maxent. *Ecography*, 40(7), 887-893.  
777

778 Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., & Ferrier, S.  
779 (2009). Sample Selection Bias and Presence-Only Distribution Models: Implications for  
780 Background and Pseudo-Absence Data. *Ecol. Appl.*, 19 (1), 181–197.  
781

782 Qiao, H., Soberón, J., & Peterson, A.T. (2015). No silver bullets in correlative ecological  
783 niche modelling: insights from testing among many potential algorithms for niche estima-  
784 tion. *Methods in Ecology and Evolution*, 6(10), 1126-1136.

785

786 Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J.,  
787 Kaschner, K., Baumgartner, M.F., Forney, K.A., Balance, L.T., Fauchald, P., Halpin, P.,  
788 Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L.F. & Werner, F.  
789 (2006). Techniques for cetacean–habitat modeling. *Marine Ecology Progress Series*, 310,  
790 271-295.

791

792 Sala, I., Harrison, C.S., & Caldeira, R.M. (2016). The role of the Azores Archipelago in  
793 capturing and retaining incoming particles. *Journal of Marine Systems*, 154, 146-156.

794

795 Scales, K. L., Hazen, E. L., Jacox, M. G., Edwards, C. A., Boustany, A. M., Oliver, M. J.,  
796 & Bograd, S. J. (2017). Scale of inference: on the sensitivity of habitat models for wide-  
797 ranging marine predators to the resolution of environmental data. *Ecography*, 40(1), 210-  
798 220.

799

800 Sillero, N. (2011) What Does Ecological Modelling Model? A Proposed Classification of  
801 Ecological Niche Models Based on Their Underlying Methods. *Ecol. Modell.*, 222, 1343–  
802 1346.

803

804 Sillero, N., Arenas-Castro, S., Enriquez-Urzelai, U., Vale, C.G., Sousa-Guedes, D.,  
805 Martínez-Freiría, F., Real, R., & Barbosa, A.M. (2021) Want to Model a Species Niche? A  
806 Step-by-Step Guideline on Correlative Ecological Niche Modelling. *Ecol. Modell.*, 456

807

808 Silva, M.A., Prieto, R., Cascão, I., Seabra, M.I., Machete, M., Baumgartner, M.F., &  
809 Santos, R.S. (2014). Spatial and temporal distribution of cetaceans in the mid-Atlantic wa-  
810 ters around the Azores. *Marine Biology Research*, 10(2), 123-137.

811

812 Schoener, T.W. (1968) The Anolis lizards of Bimini: resource partitioning in a complex  
813 fauna. *Ecology*, 49, 704–726.

814

815 Scott, M.D., Chivers, S.J., Olson, R.J., & Lindsay, R.J. (1993). Radiotracking of spotted  
816 dolphins associated with tuna in the eastern tropical Pacific. In Abstracts of the 10th Bien-  
817 nial Conference on the Biology of Marine Mammals, Galveston, Texas, 11–15 November  
818 1993. Society for Marine Mammalogy, Lawrence, Kans. p. 97.

819

820 Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: concepts, methods,  
821 and assumptions. *Proceedings of the National Academy of Sciences*, 106(Supplement 2),  
822 19644-19650.

823 Soberon, J. & Peterson, T. (2004). Biodiversity informatics: managing and applying  
824 primary biodiversity data. *Phil. Trans. R. Soc. B* 359: 689 698.

825 Steele, J. H. (1991). Marine ecosystem dynamics: comparison of scales. *Ecol. Res.*, 6: 175-  
826 183.

827



828 Tepsich, P., Rosso, M., Halpin, P.N., & Moulins, A. (2014). Habitat preferences of two  
829 deep-diving cetacean species in the northern Ligurian Sea. *Marine Ecology Progress*  
830 *Series, 508*, 247-260.

831

832 Virgili, A., Racine, M., Authier, M., Monestiez, P. & Ridoux, V. (2017). Comparison of  
833 habitat models for scarcely detected species. *Ecological Modelling, 346*, 88-98.

834

835 Warren, D., Dinnage, R., Matzke, N. & Cardillo, M. (2017). ENMTools: Analysis of niche  
836 evolution using niche and distribution models. R package version 0.2.

837

838 Warren, D.L., Wright, A.N., Seifert, S.N. & Shaffer, H.B. (2014). Incorporating model  
839 complexity and spatial sampling bias into ecological niche models of climate change risks  
840 faced by 90 California vertebrate species of concern. *Diversity and Distributions, 20*, 334-  
841 343

842

843 Warren, D. L., Beaumont, L. J., Dinnage, R., & Baumgartner, J. (2019). New methods for  
844 measuring ENM breadth and overlap in environmental space. *Ecography, 42*(3), 444-446.

845

846 Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J. E., & Willis, K.J. (2005).  
847 Conservation biogeography: assessment and prospect. *Diversity and distributions, 11*(1), 3-  
848 23.

849

850 Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood es-  
851 timation of semiparametric generalized linear models. *Journal of the Royal Statistical Soci-*  
852 *ety (B)* 73(1):3-36

853

854

855

856

857

858

859

860

*Supplemental material: Appendix S1*

861

862 **To be or not to be: the role of absences in niche modelling for highly**

863 **motile species in dynamic marine environments**

864

865 Fernandez M<sup>1,2</sup>, Sillero N<sup>3</sup>, Yesson C<sup>4</sup>

866

867 1. MARE - Marine and Environmental Sciences Centre, Agência Regional para o Desenvolvimento

868 da Investigação Tecnologia e Inovação (ARDITI), Funchal, Portugal

869

870 2. cE3c/Azorean Biodiversity Group, Departamento de Biologia, Faculdade de Ciências e

871 Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal

872

873 3. CICGE - Centro de Investigação em Ciências Geo-Espaciais, Alameda do Monte da Virgem,

874 4430-146 Vila Nova de Gaia, Portugal

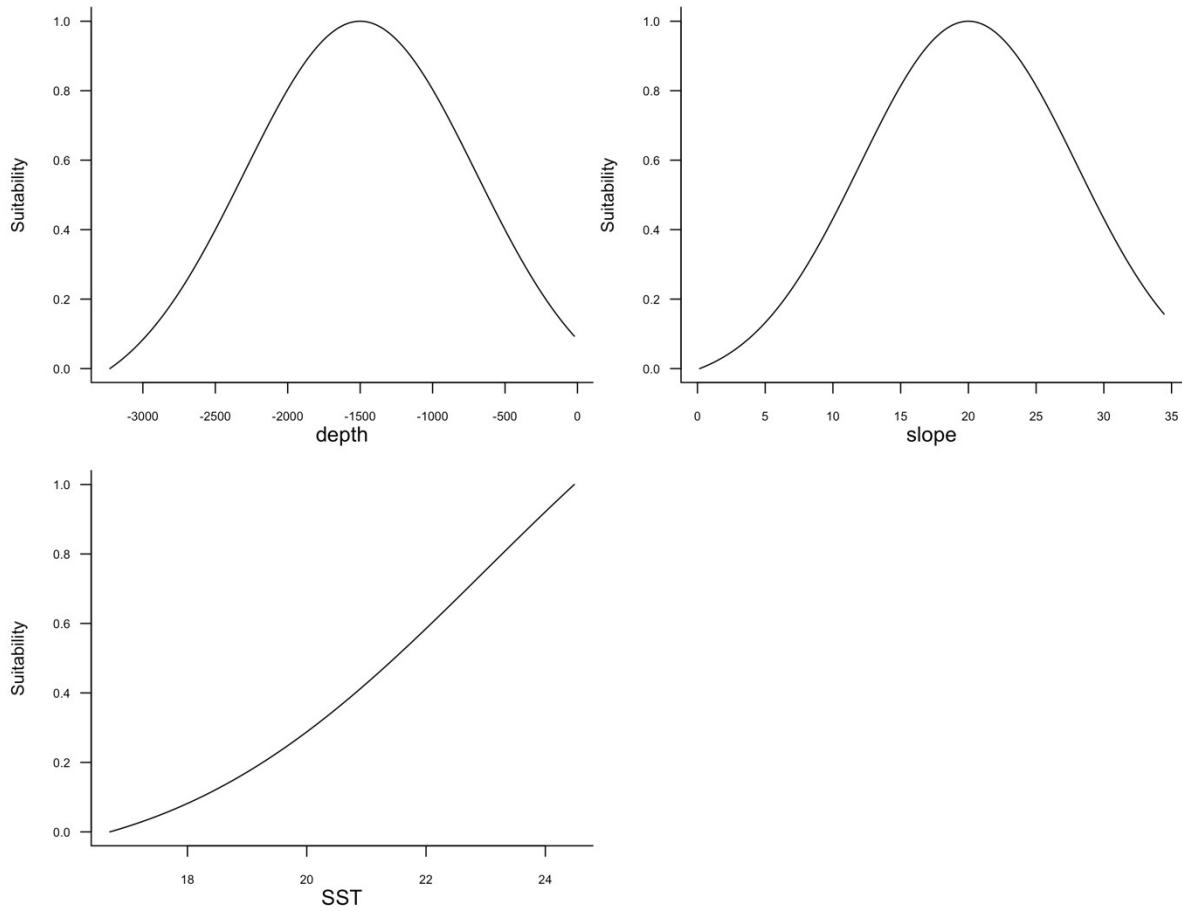
875

876 4. Zoological Society of London, Institute of Zoology, London, UK

877

878

879



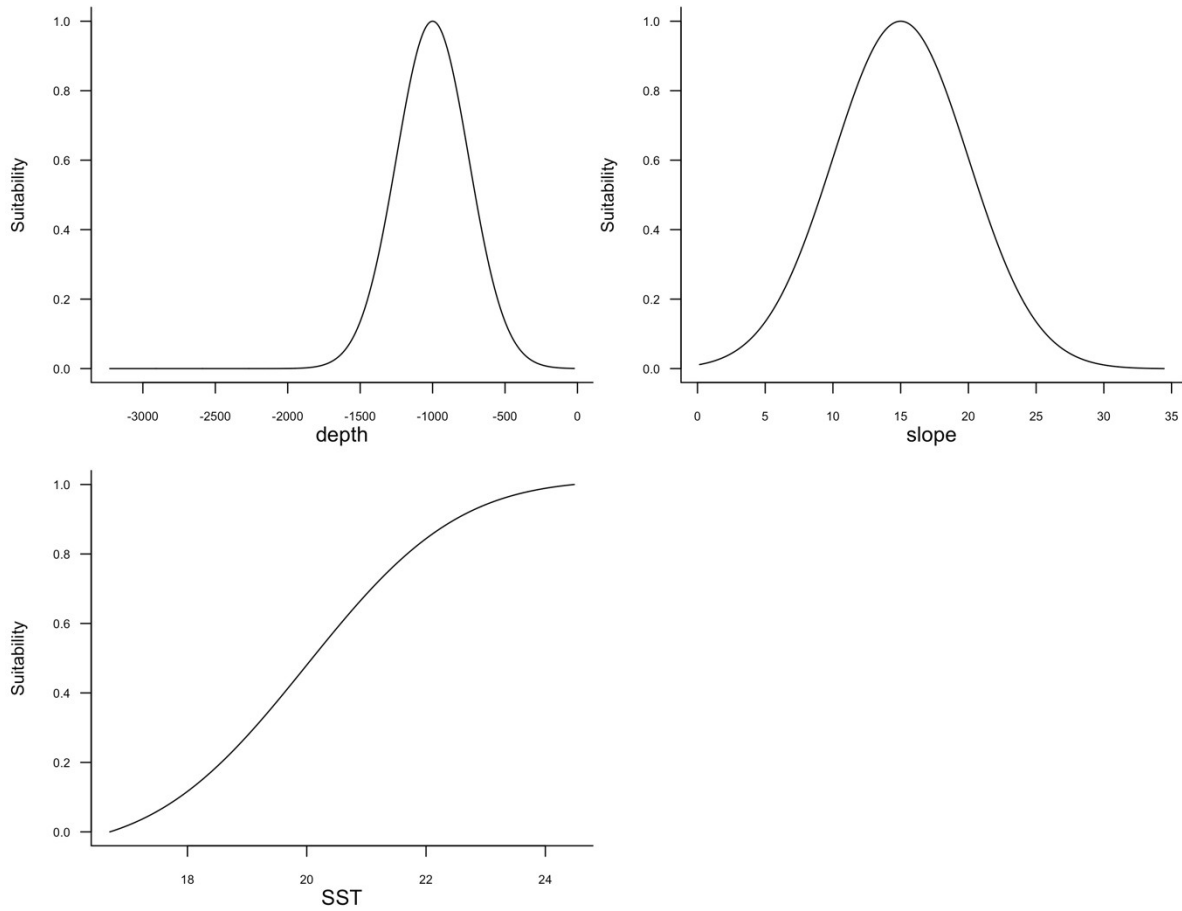
880

881 Figure S1: Response curves for the generalist static species, depicting the relation of the

882 suitability with the environmental variables used to build the virtual species.

883

884

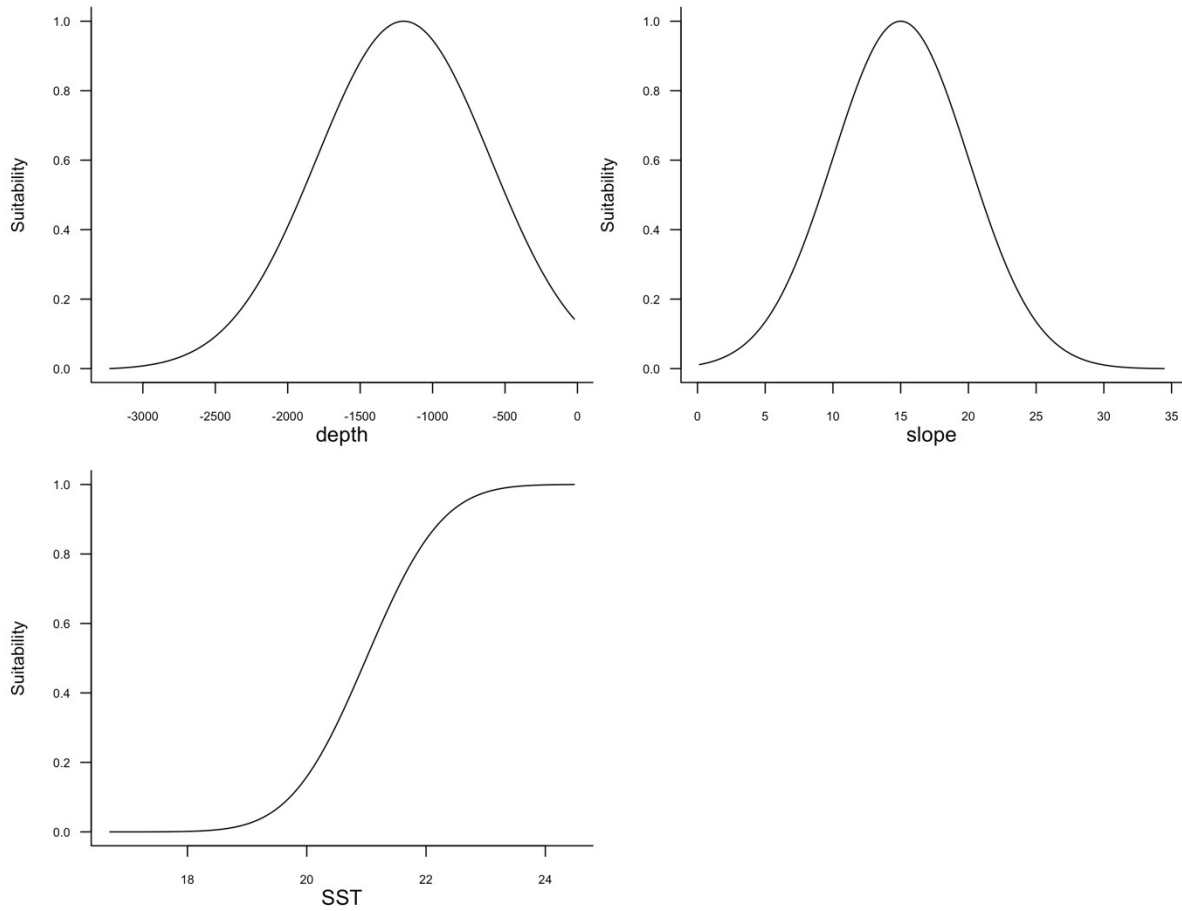


885

886 Figure S2: Response curves for the specialist static species, depicting the relation of the

887 suitability with the environmental variables used to build the virtual species.

888

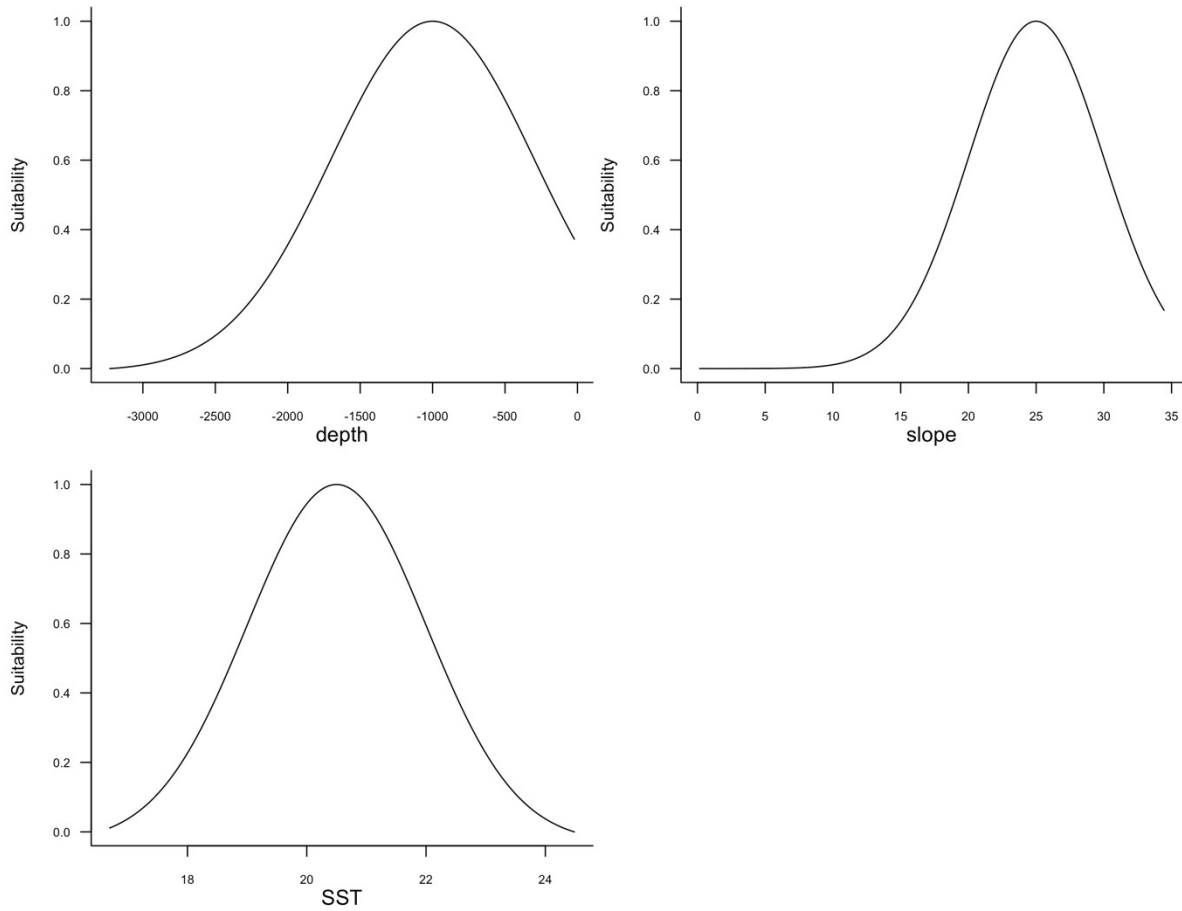


889

890 Figure S3: Response curves for the generalist dynamic species, depicting the relation of the

891 suitability with the environmental variables used to build the virtual species.

892



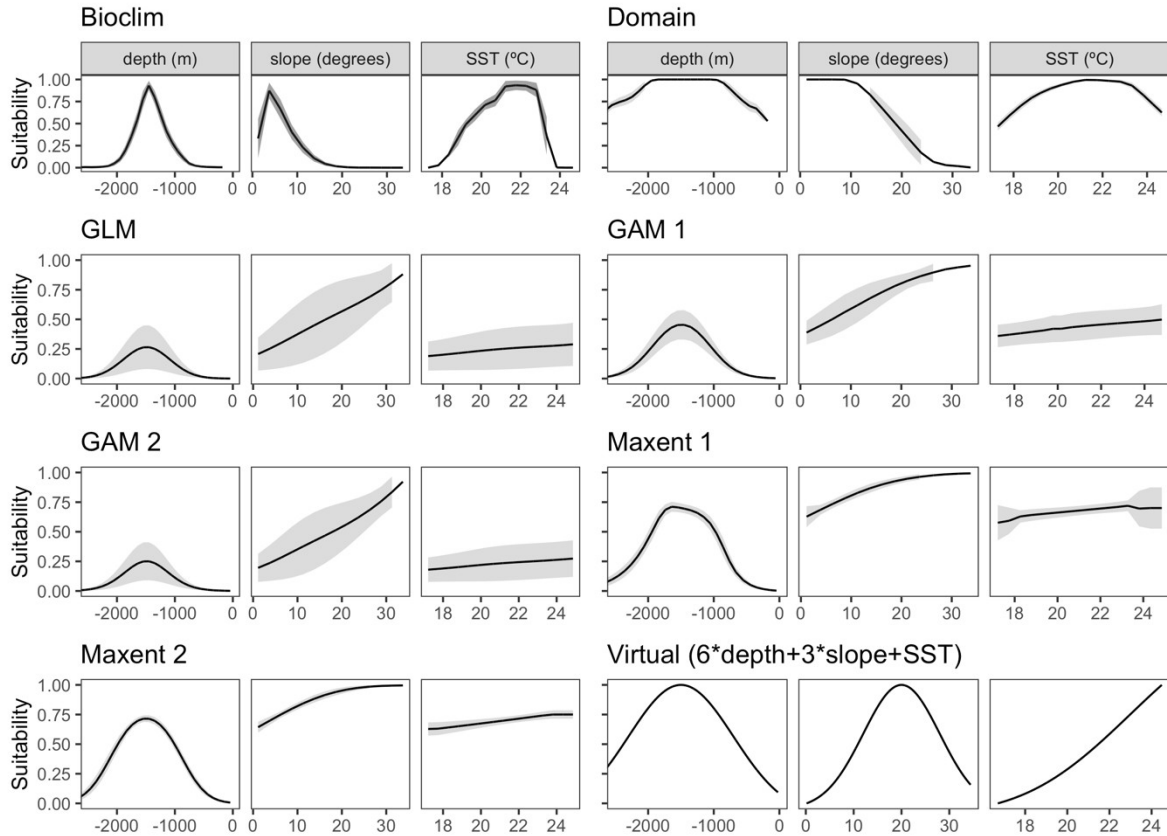
893

894 Figure S4: Response curves for the specialist dynamic species, depicting the relation of the

895 suitability with the environmental variables used to build the virtual species.

896

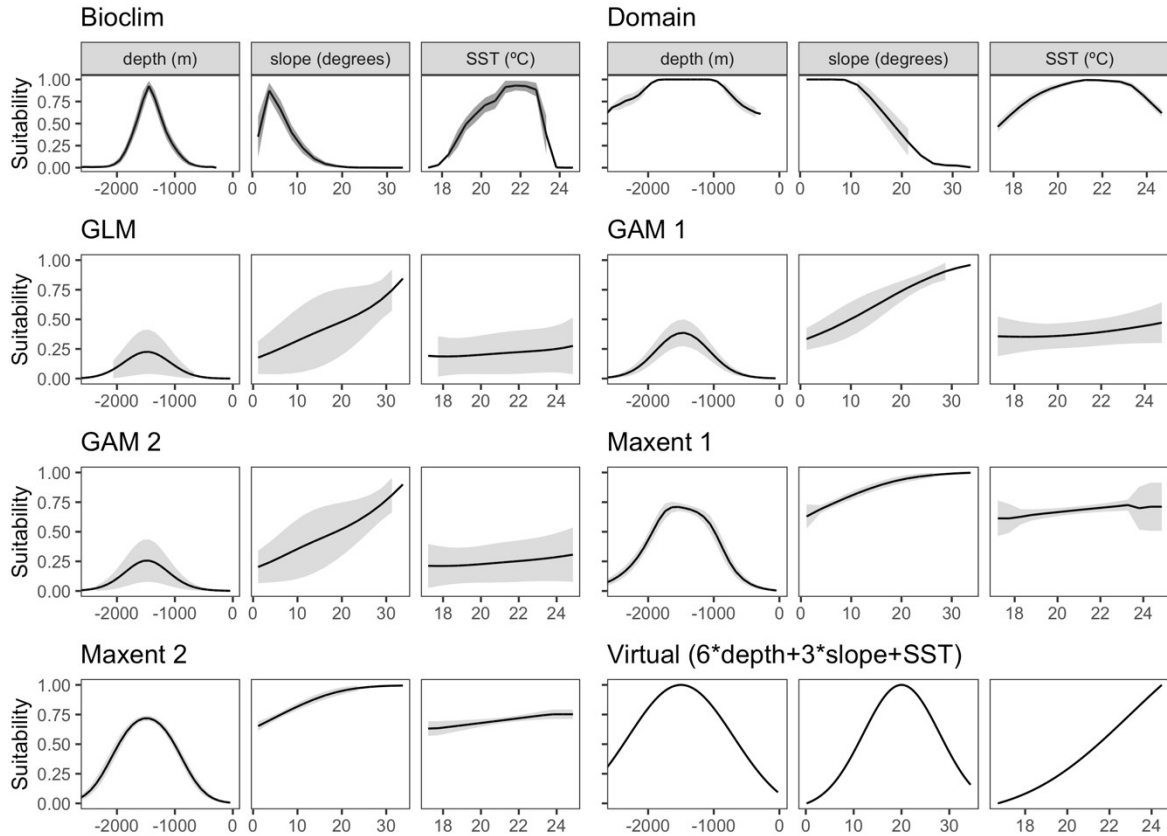
897



898  
 899  
 900  
 901  
 902  
 903

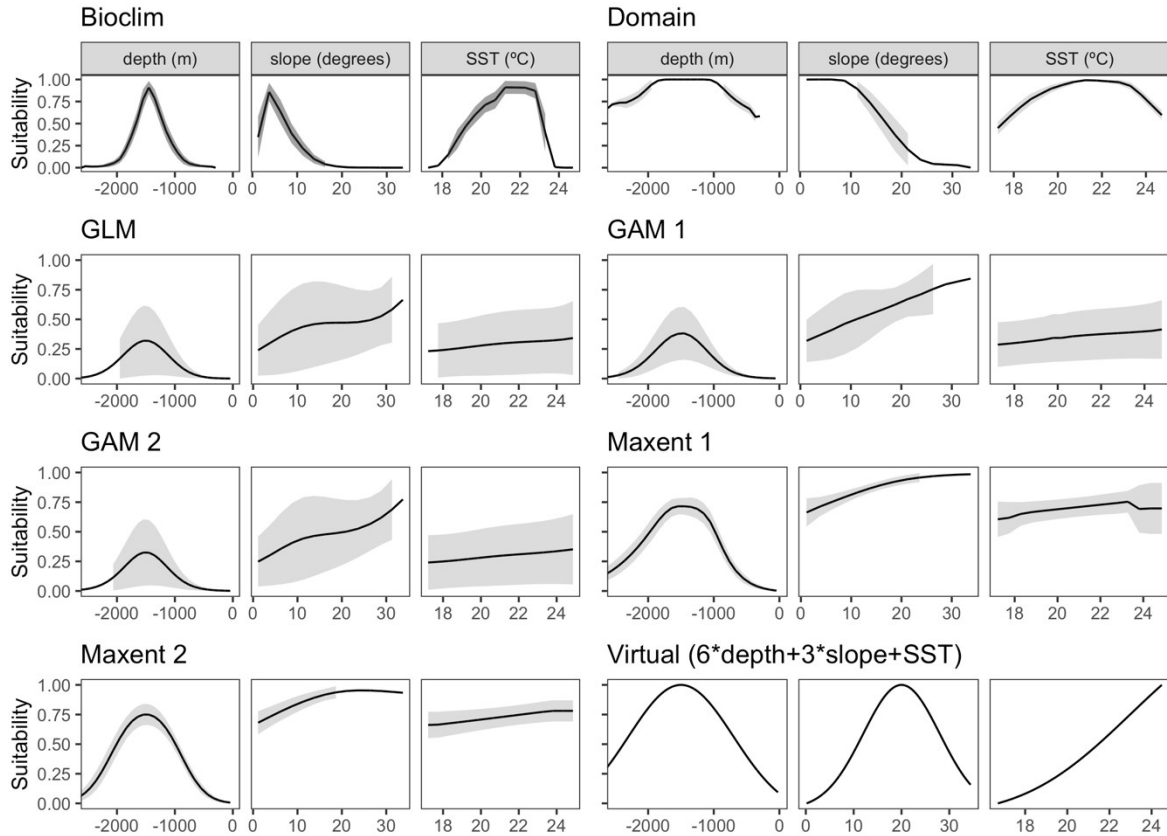
Figure S5: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist static species with encounter rate 1.4 groups/100km.





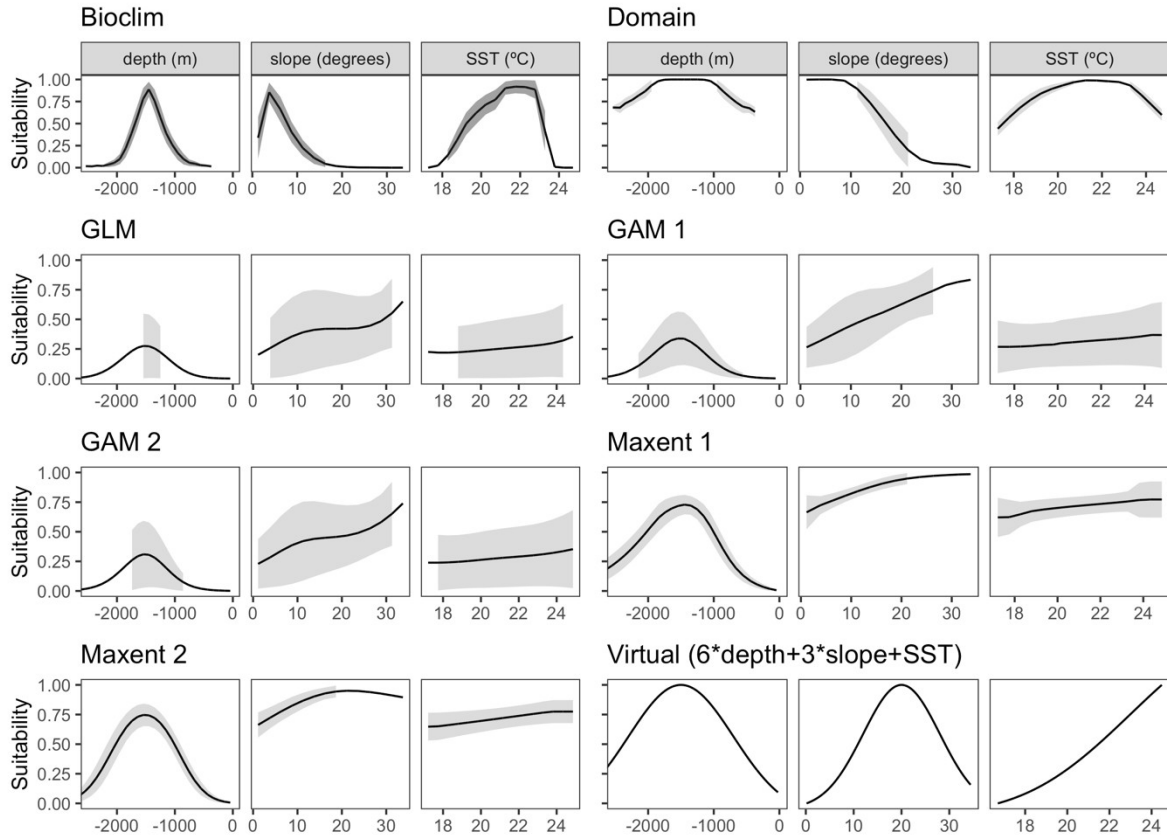
904  
 905  
 906  
 907  
 908

Figure S6: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist static species with encounter rate 1 group/100km.



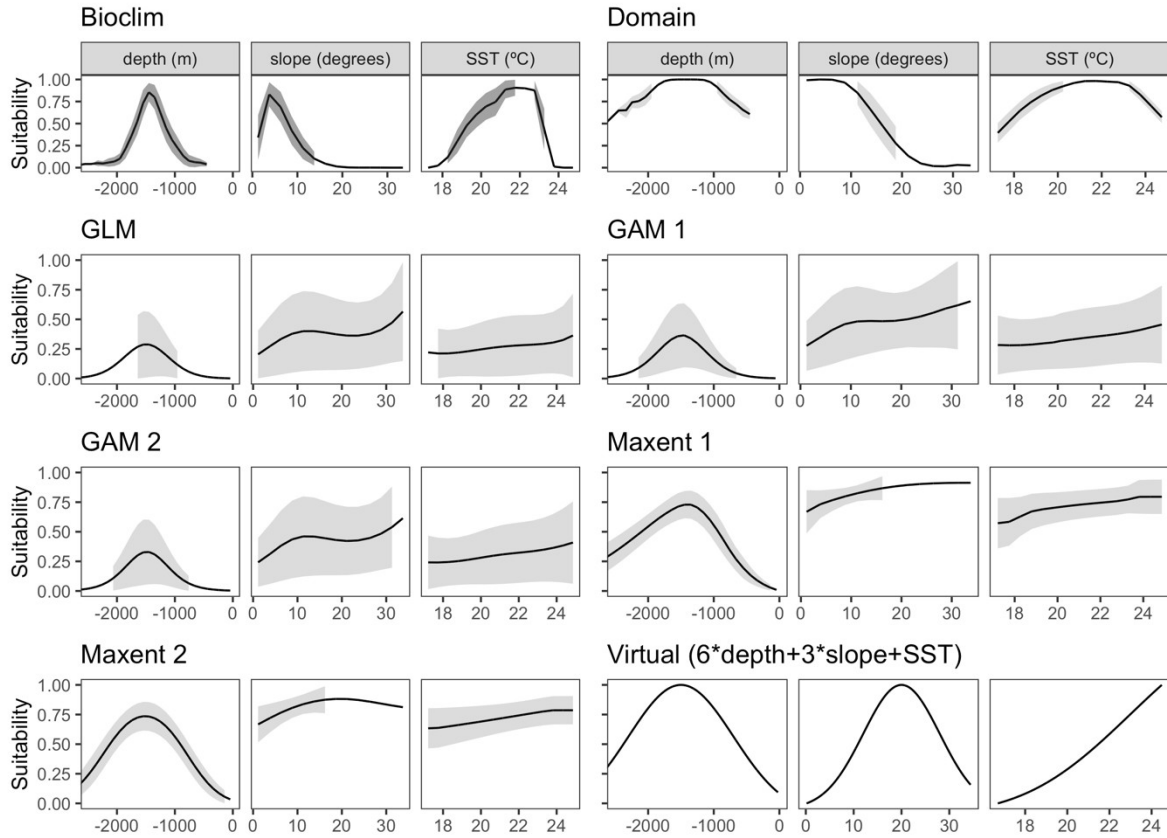
909  
 910  
 911  
 912  
 913

Figure S7: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist static species with encounter rate 0.8 groups/100km.



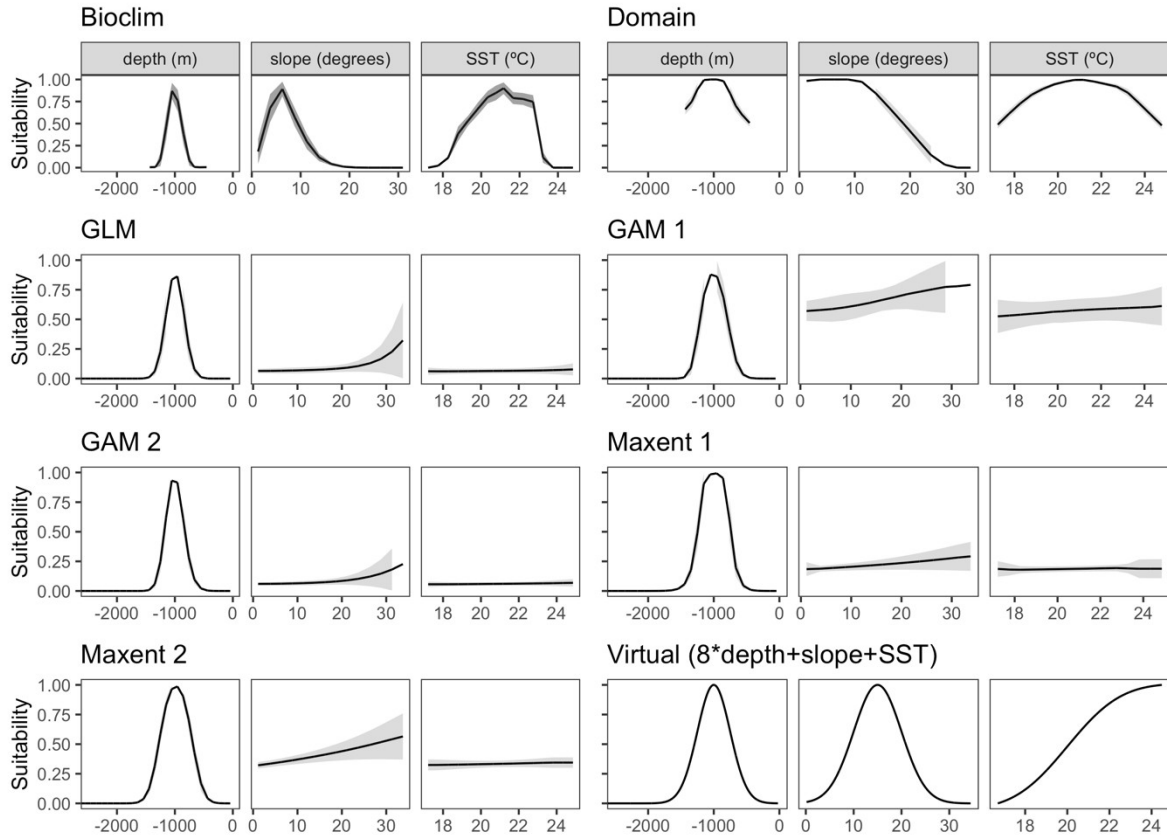
914  
 915  
 916  
 917  
 918

Figure S8: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist static species with encounter rate 0.6 groups/100km.



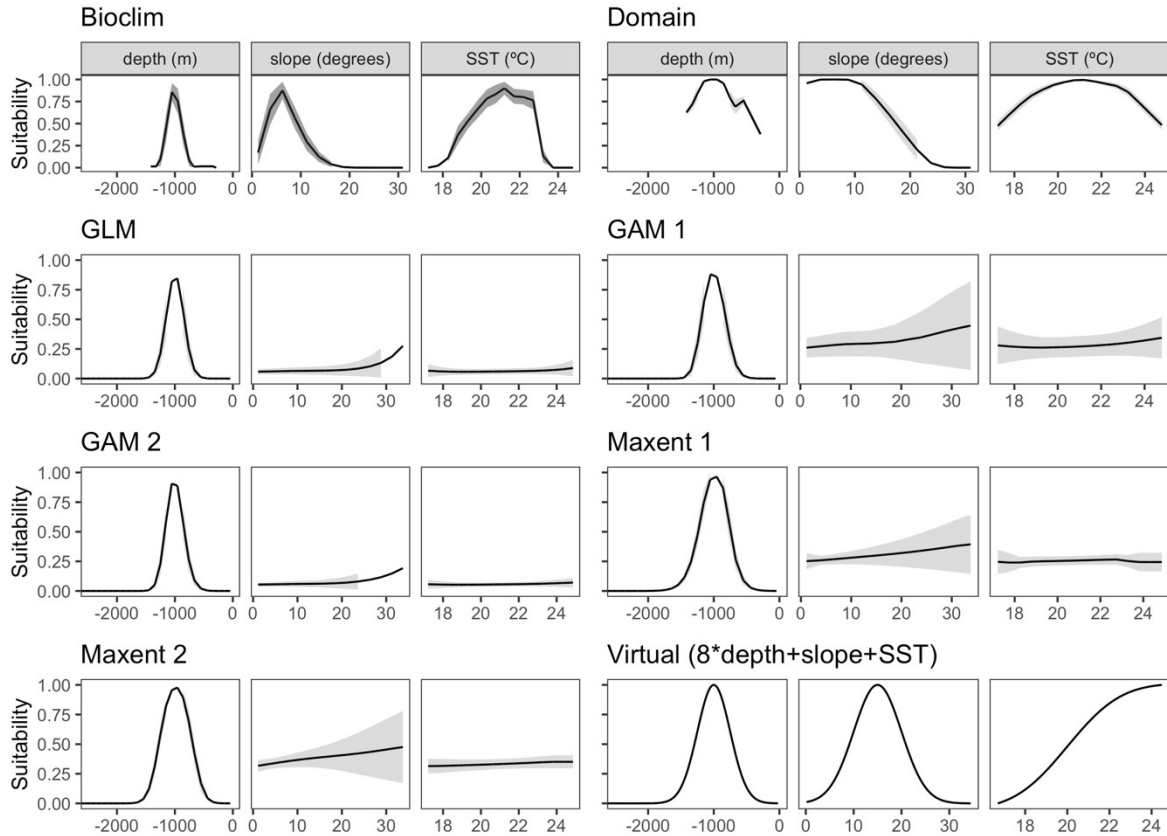
919  
 920  
 921  
 922  
 923

Figure S9: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist static species with encounter rate 0.2 groups/100km.



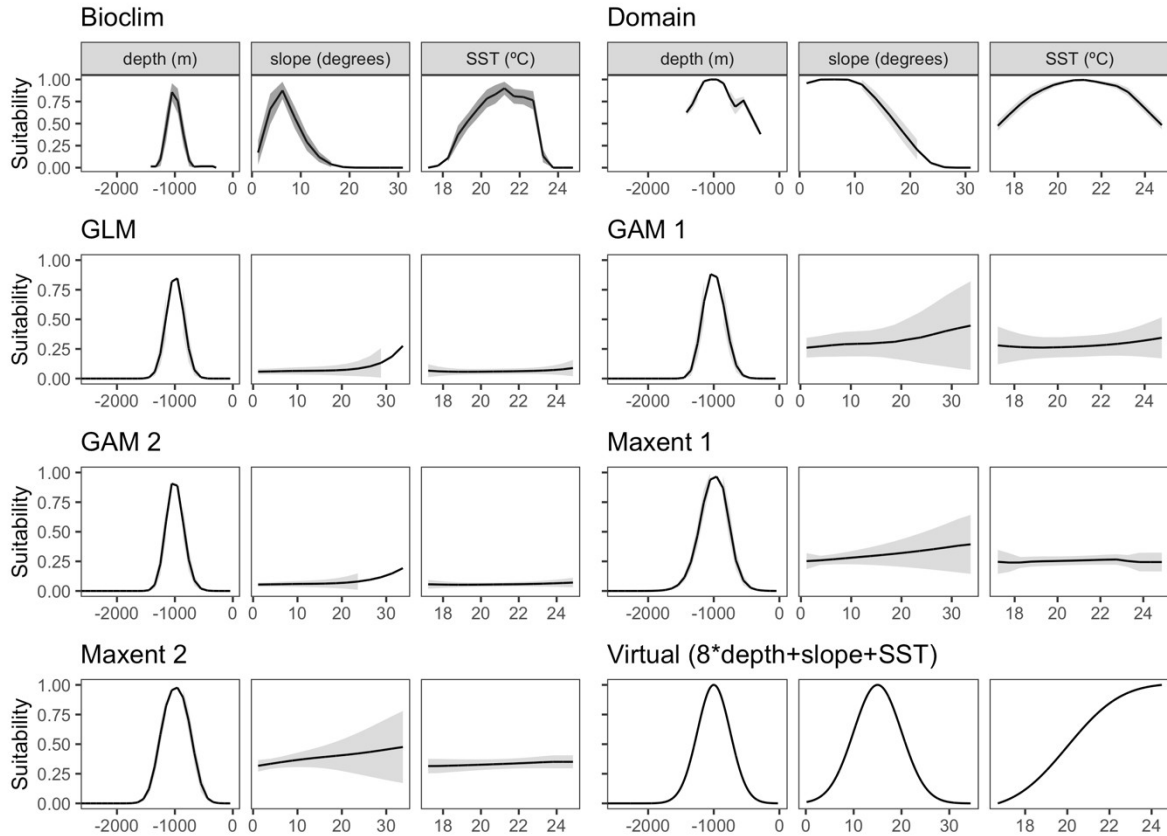
924  
 925  
 926  
 927  
 928

Figure S10: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the specialist static species with encounter rate 1.4 groups/100km.

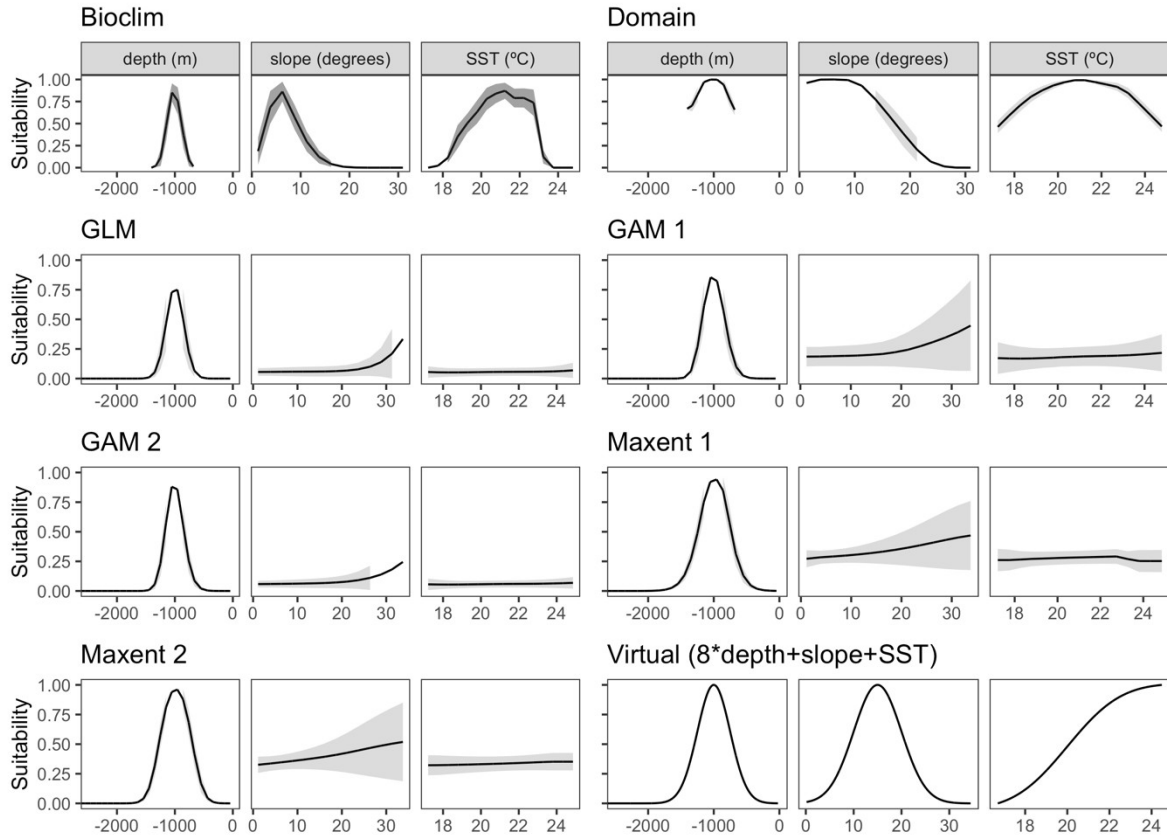


929  
 930  
 931  
 932  
 933

Figure S11: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the specialist static species with encounter rate 1 group/100km.



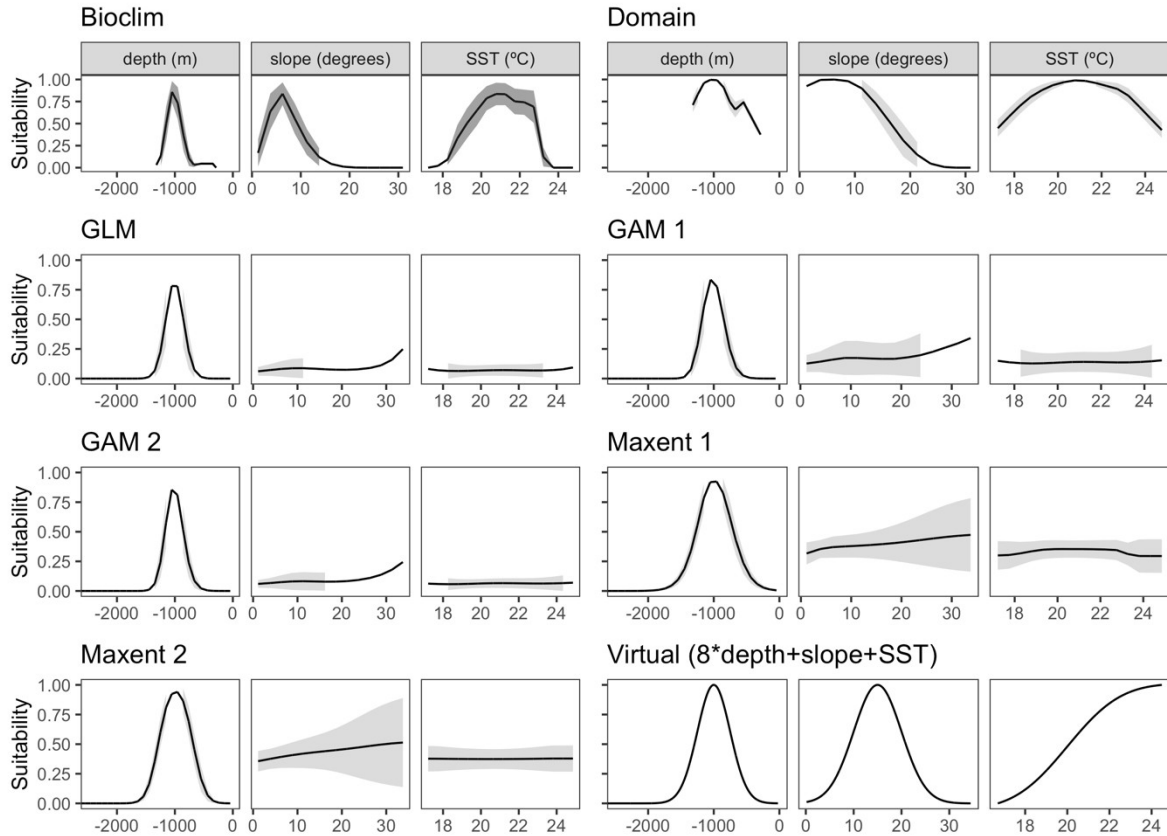
934  
 935 Figure S12: Response curves outputs from the different modelling algorithms, together  
 936 with the theoretical curves (at the bottom right) regarding the specialist static species with  
 937 encounter rate 0.8 groups/100km.  
 938



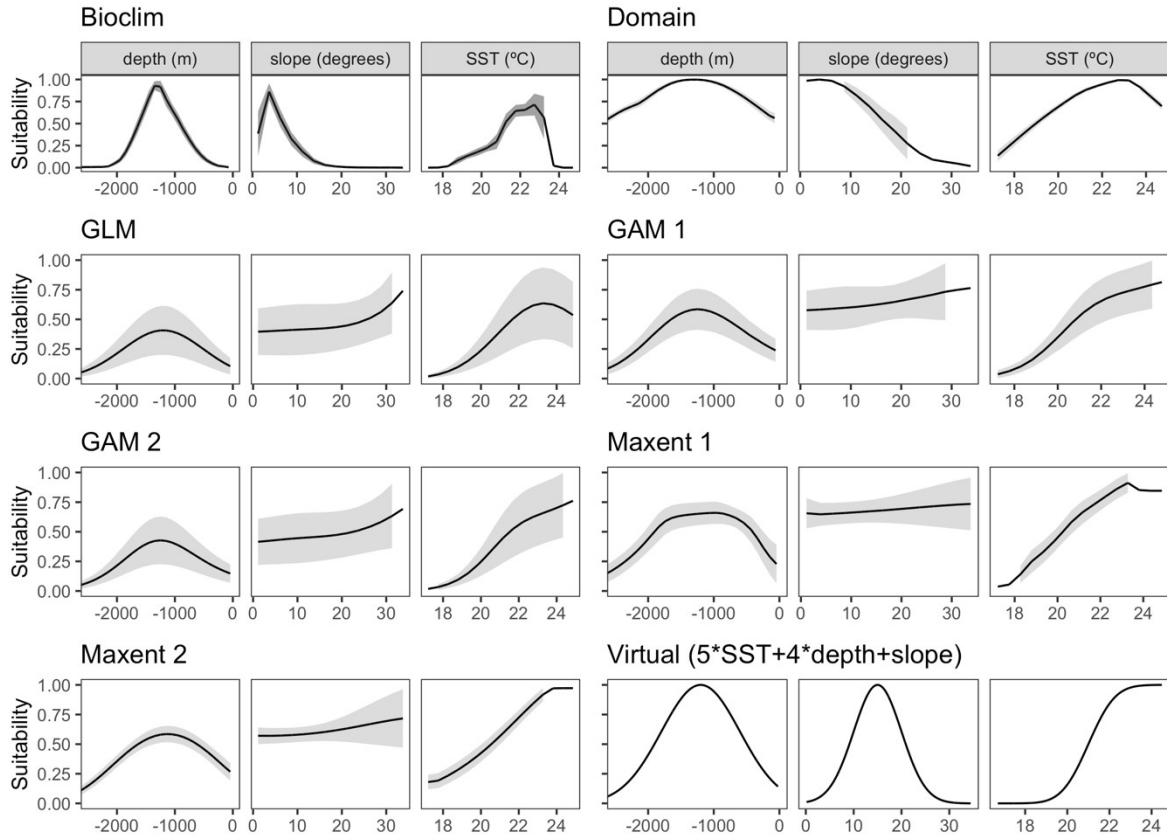
939  
 940  
 941  
 942  
 943  
 944

Figure S13: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the specialist static species with encounter rate 0.6 groups/100km.



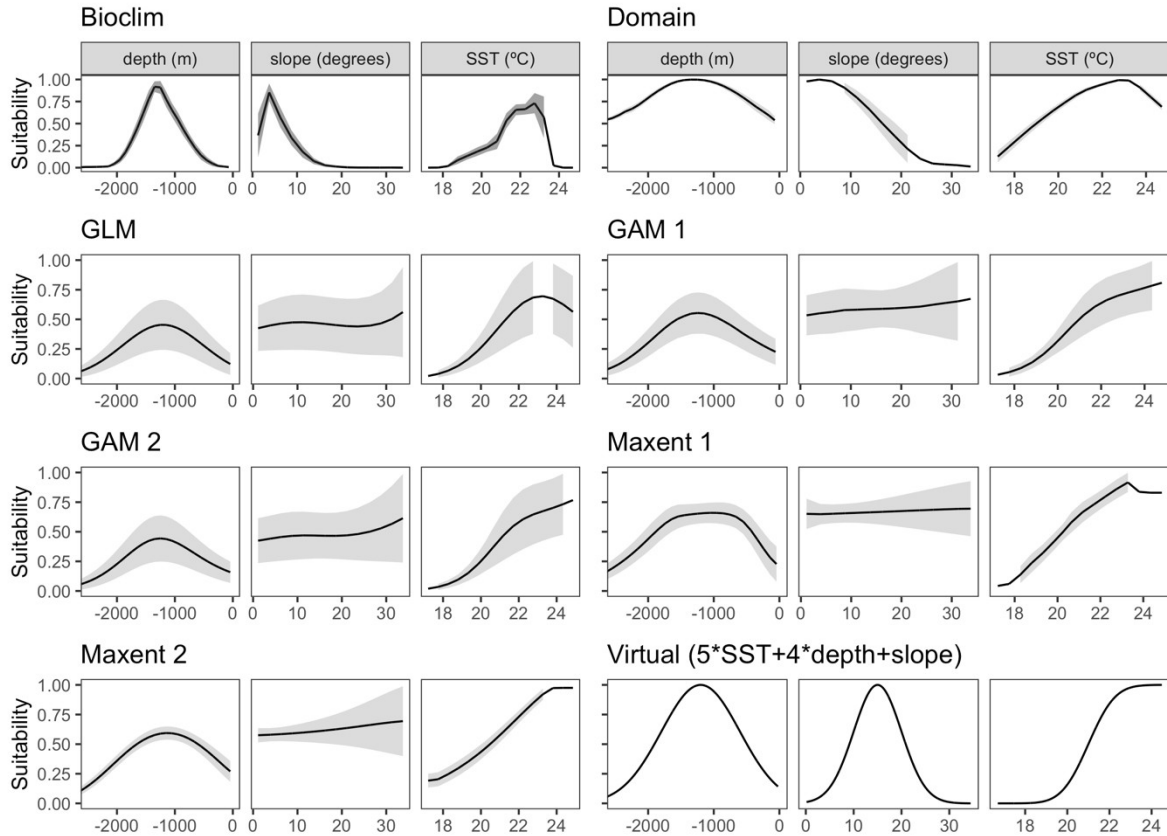


945  
 946 Figure S14: Response curves outputs from the different modelling algorithms, together  
 947 with the theoretical curves (at the bottom right) regarding the specialist static species with  
 948 encounter rate 0.2 groups/100km.  
 949



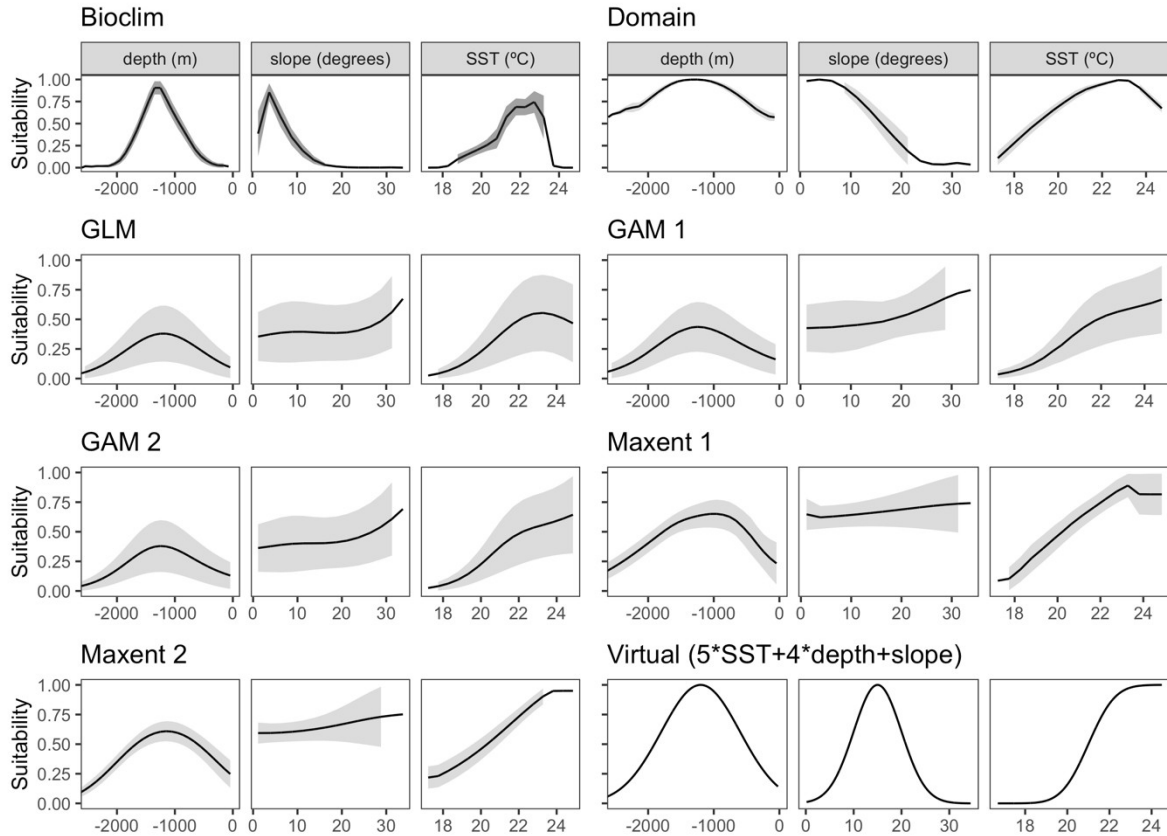
950  
 951  
 952  
 953  
 954

Figure S15: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist dynamic species with encounter rate 1.4 groups/100km.

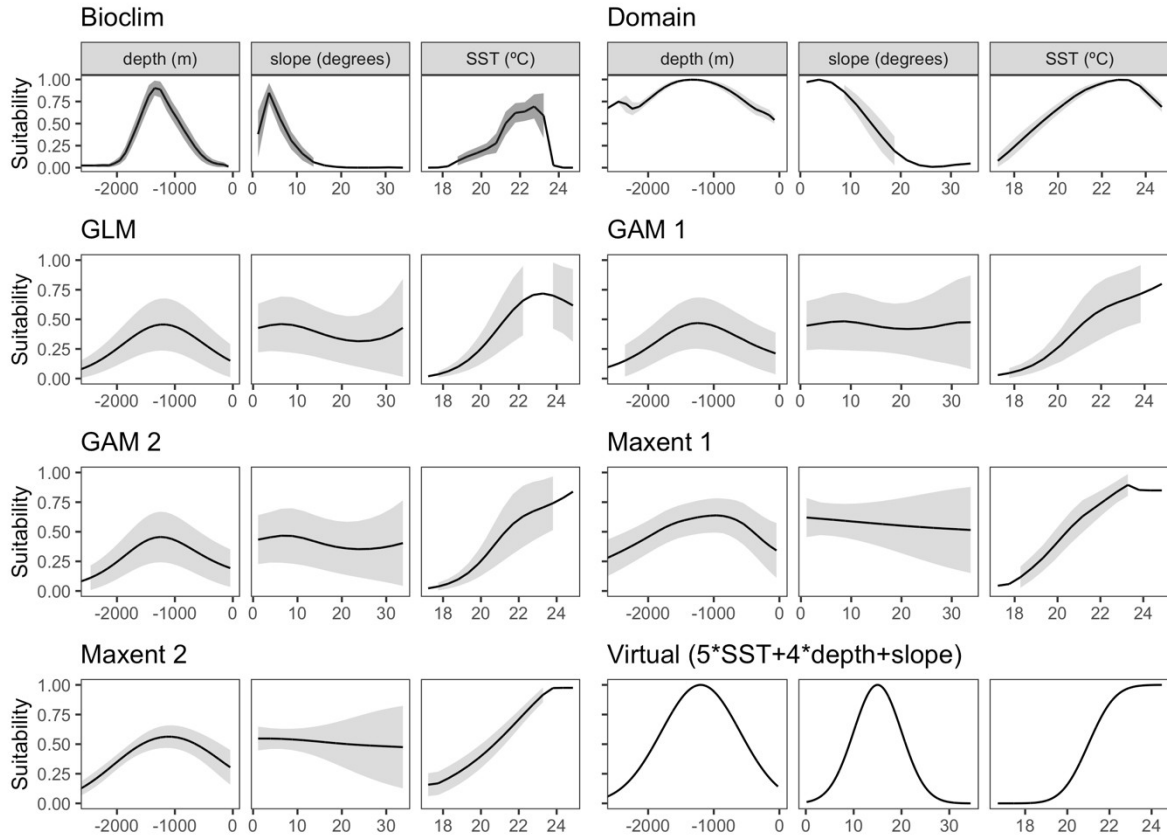


955  
 956  
 957  
 958  
 959  
 960

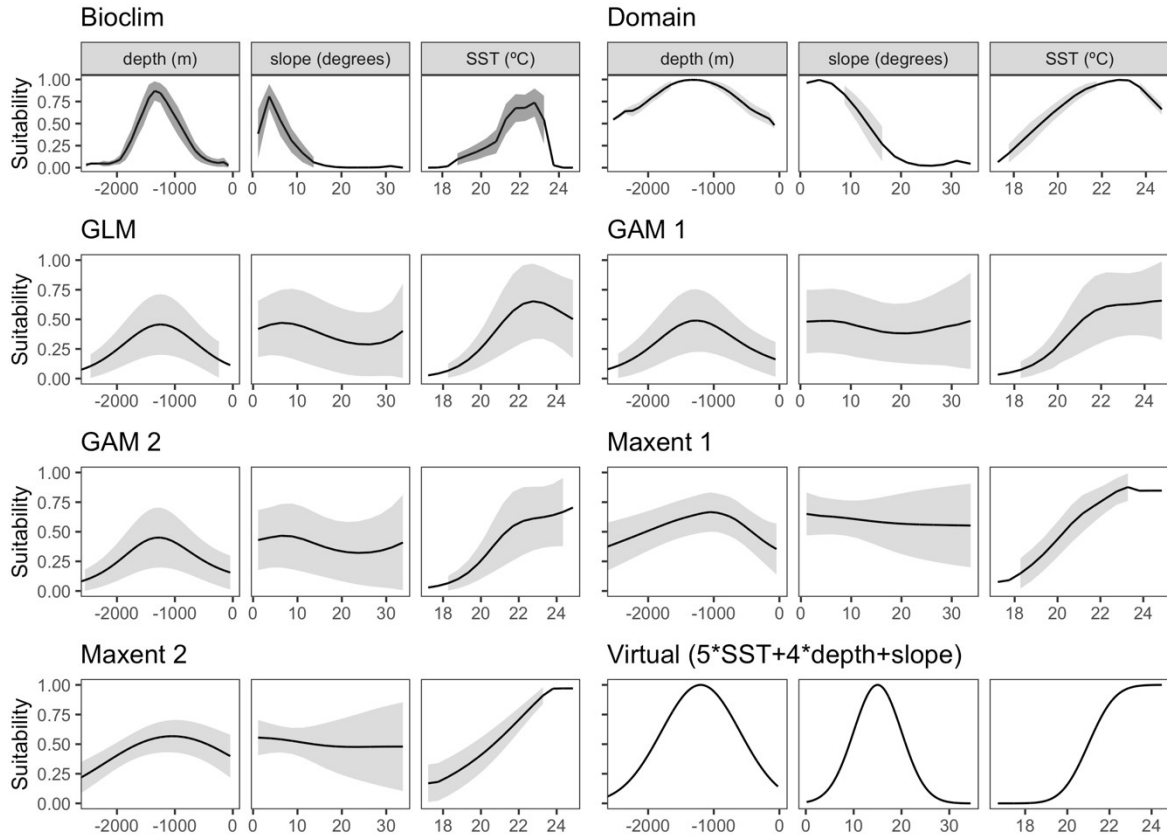
Figure S16: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist dynamic species with encounter rate 1 group/100km.



961  
 962 Figure S17: Response curves outputs from the different modelling algorithms, together  
 963 with the theoretical curves (at the bottom right) regarding the generalist dynamic species  
 964 with encounter rate 0.8 groups/100km.  
 965

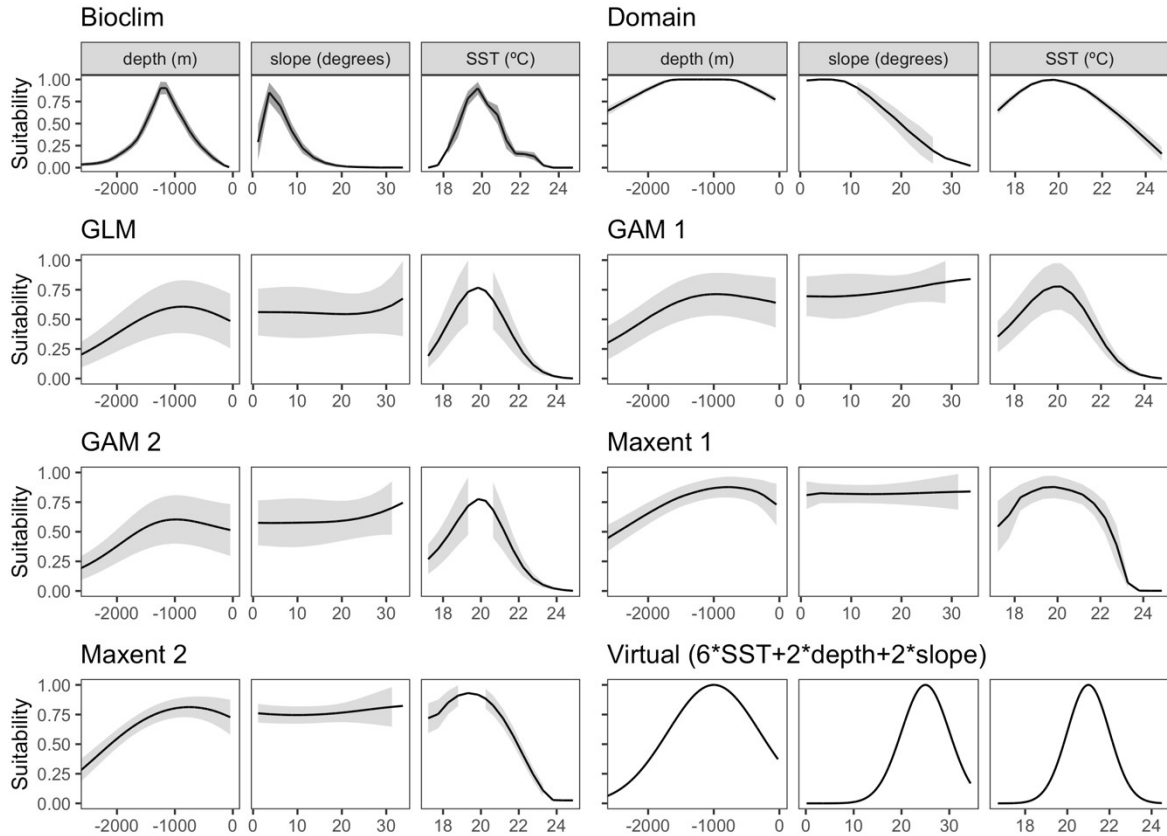


966  
 967 Figure S18: Response curves outputs from the different modelling algorithms, together  
 968 with the theoretical curves (at the bottom right) regarding the generalist dynamic species  
 969 with encounter rate 0.6 groups/100km.  
 970



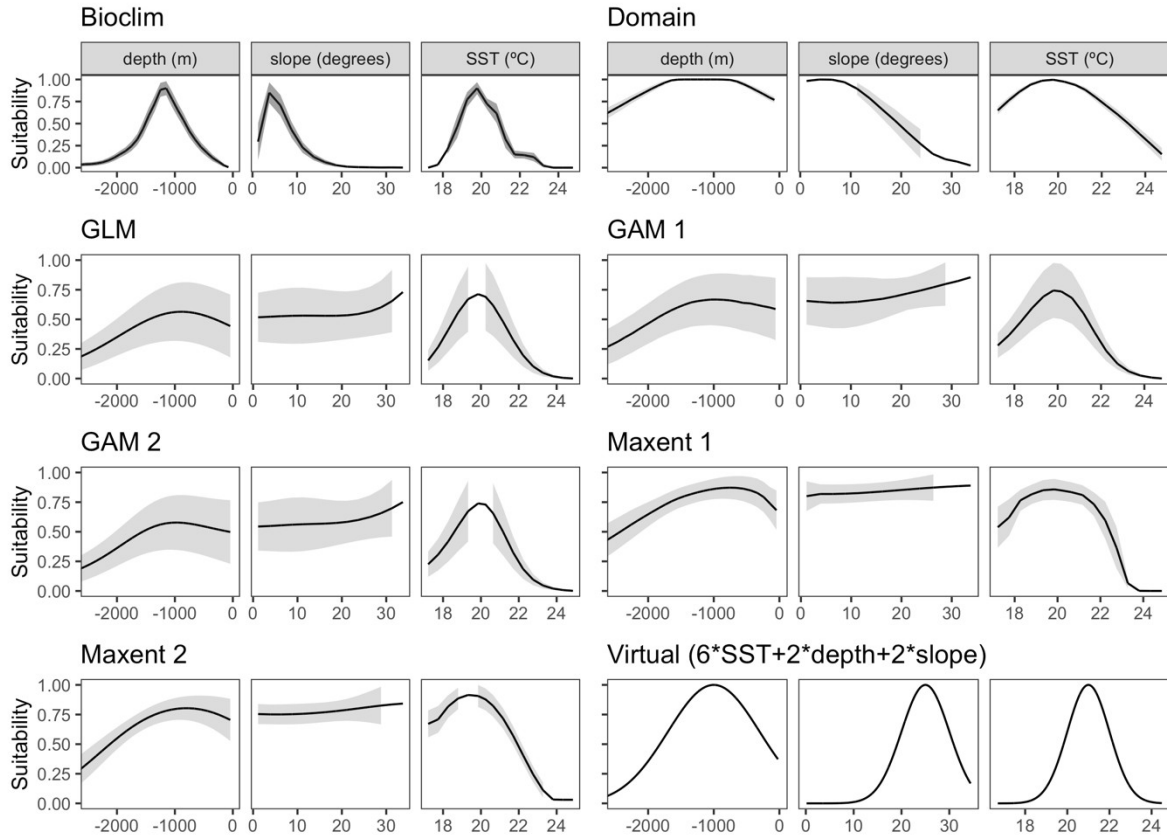
971  
 972  
 973  
 974  
 975

Figure S19: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist dynamic species with encounter rate 0.2 groups/100km.



976  
 977  
 978  
 979  
 980

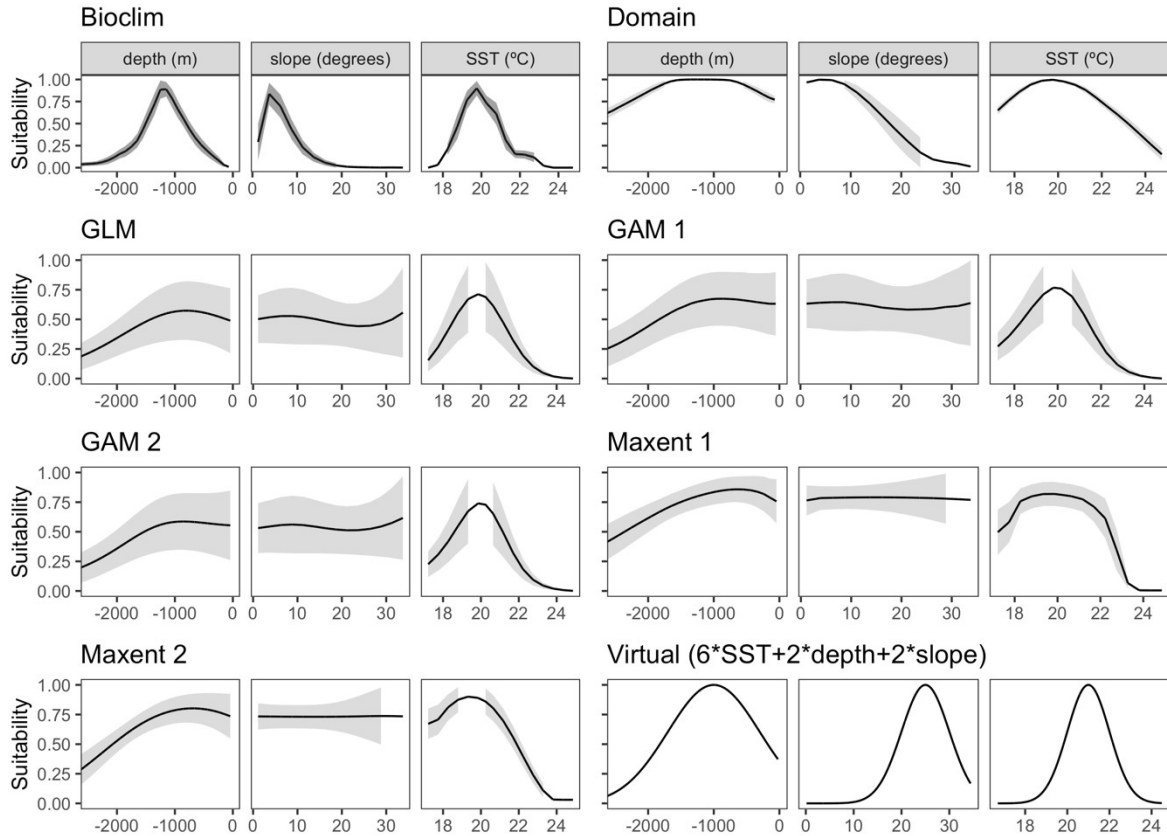
Figure S20: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist dynamic species with encounter rate 1.4 groups/100km.



981  
 982  
 983  
 984  
 985

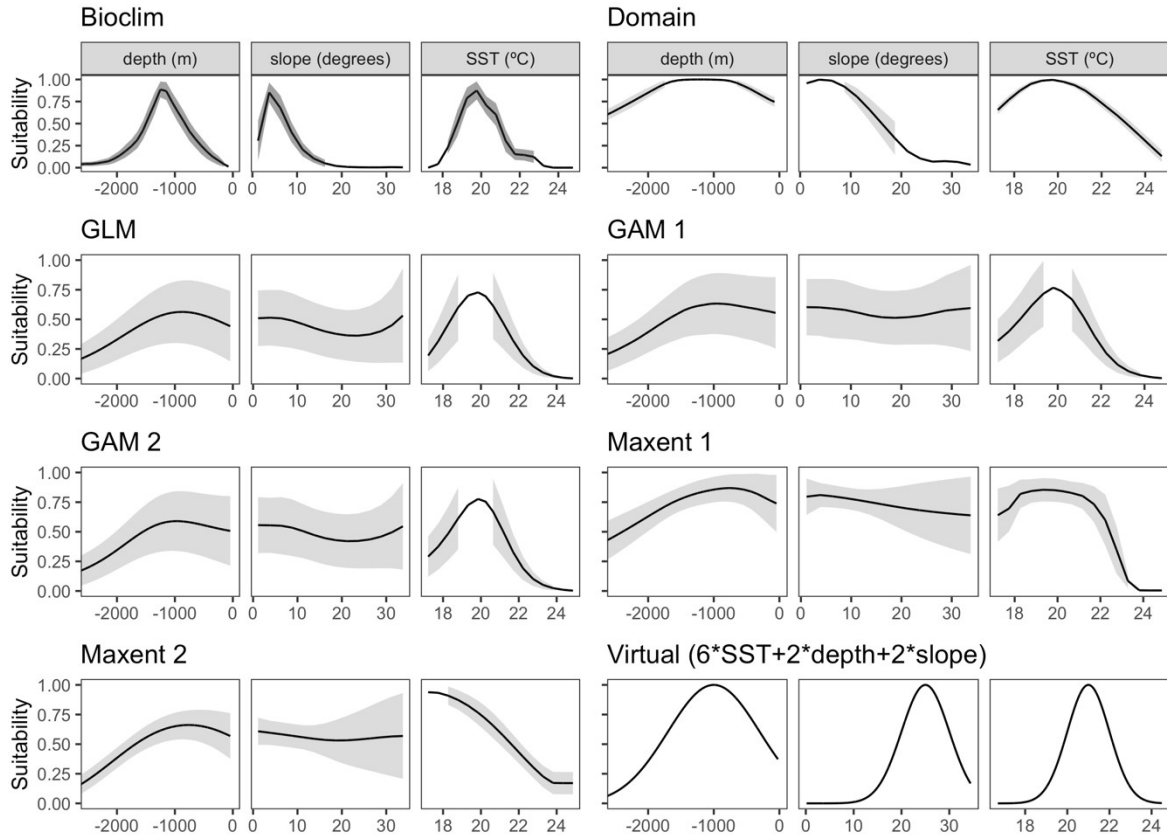
Figure S21: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist dynamic species with encounter rate 1 group/100km.



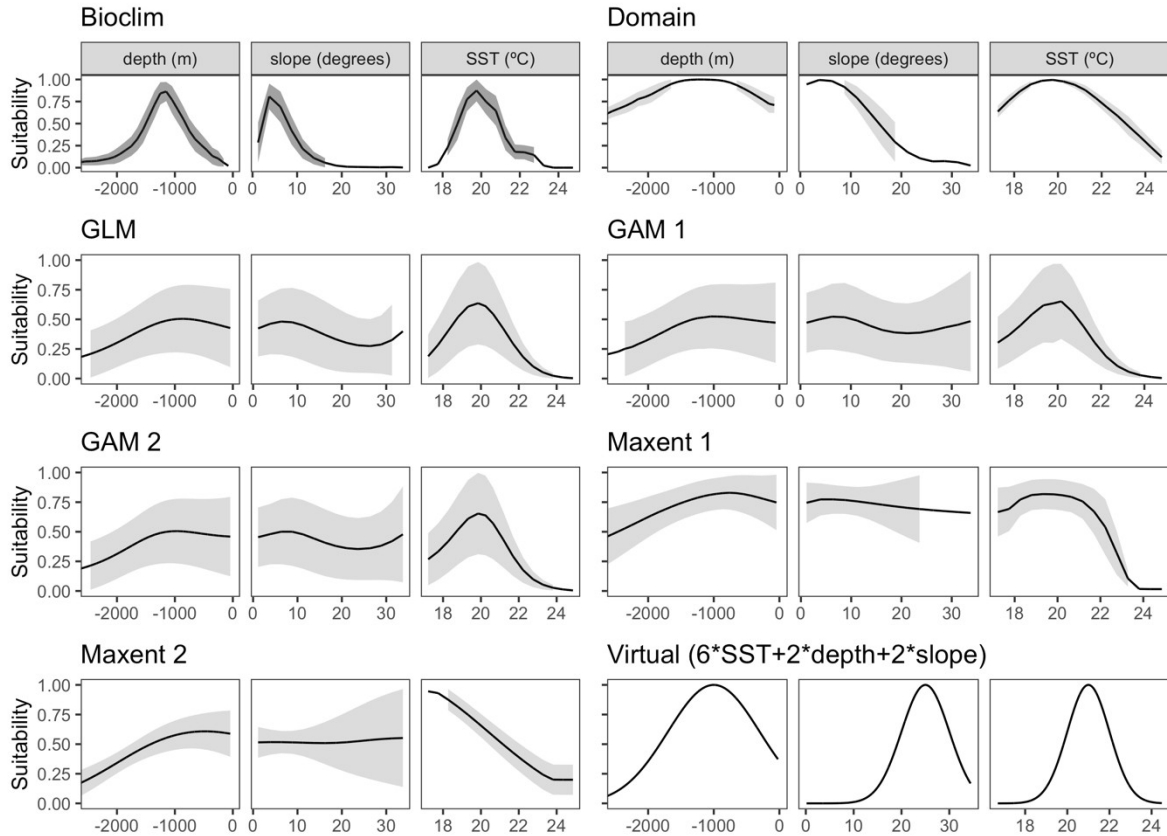


986  
 987  
 988  
 989  
 990

Figure S22: Response curves outputs from the different modelling algorithms, together with the theoretical curves (at the bottom right) regarding the generalist dynamic species with encounter rate 0.8 groups/100km.



991  
 992 Figure S23: Response curves outputs from the different modelling algorithms, together  
 993 with the theoretical curves (at the bottom right) regarding the generalist dynamic species  
 994 with encounter rate 0.6 groups/100km.  
 995



996  
 997 Figure S24: Response curves outputs from the different modelling algorithms, together  
 998 with the theoretical curves (at the bottom right) regarding the generalist dynamic species  
 999 with encounter rate 0.2 groups/100km.the

1000  
 1001  
 1002  
 1003  
 1004  
 1005  
 1006  
 1007

1008

1009

***Supplemental material: Appendix S2***

1010

***ODMAP table***

1011

1012

**To be or not to be: the role of absences in niche modelling for highly**

1013

**motile species in dynamic marine environments**

1014

1015

Fernandez M<sup>1,2</sup>, Sillero N<sup>3</sup>, Yesson C<sup>4</sup>

1016

1017

1. MARE - Marine and Environmental Sciences Centre, Agência Regional para o Desenvolvimento

1018

da Investigação Tecnologia e Inovação (ARDITI), Funchal, Portugal

1019

1020

2. cE3c/Azorean Biodiversity Group, Departamento de Biologia, Faculdade de Ciências e

1021

Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal

1022

1023

3. CICGE - Centro de Investigação em Ciências Geo-Espaciais, Alameda do Monte da Virgem,

1024

4430-146 Vila Nova de Gaia, Portugal

1025

1026

4. Zoological Society of London, Institute of Zoology, London, UK

1027

<b>ODMAP element</b>	<b>Content</b>
<b>OVERVIEW</b>	
Authorship	<ul style="list-style-type: none"> <li>• <b>Authors:</b> Fernandez, M; Sillero, N; Yesson, C.</li> <li>• <b>Contact e-mail:</b> marc.fern@gmail.com</li> <li>• <b>Title:</b> To be or not to be: the role of absences in niche modelling for highly mobile species in dynamic marine environments</li> </ul>
Model objective	<ul style="list-style-type: none"> <li>• <b>SDM objective/purpose:</b> ecological inference / explanation</li> <li>• <b>Main target output:</b> continuous habitat suitability index</li> </ul>
Taxon	Highly dynamic marine pelagic species (e.g. sharks, tuna, cetaceans...)
Location	Azores archipelago, Portugal.
Scale of analysis	<ul style="list-style-type: none"> <li>• <b>Spatial Extent (Lon / Lat):</b> Longitude 32°W-24°W, Latitude 36°N-40°N</li> <li>• <b>Spatial resolution:</b> 2km</li> <li>• <b>Temporal extent/time period:</b> May to August (2013-2016)</li> <li>• <b>Temporal resolution:</b> 8-days</li> <li>• <b>Type of extent boundary:</b> rectangular</li> </ul>
Biodiversity data overview	<ul style="list-style-type: none"> <li>• <b>Observation type:</b> Simulated transect data</li> <li>• <b>Response/data type:</b> Presence-absence</li> </ul>
Type of predictors	Climatic, topographic
Conceptual model	<p><b>Hypotheses about species-environment relationships:</b> Using a set of 20 virtual species and 3 environmental variables we tested how the presence of methodological absences in simulated standardized surveys might influence the accuracy of the predictions when using different modelling techniques.</p>
Assumptions	Species–environment equilibrium, Availability of all important predictors, Perfect detectability
SDM algorithms	• <b>Model algorithms:</b> BIOCLIM, DOMAIN, GLM, GAM, MAXENT

	<ul style="list-style-type: none"> <li>• <b>Justification of model complexity:</b> BIOCLIM and DOMAIN were fitted using the default settings. For the GLM quadratic approach was applied to allow more flexibility on the responses. Two different approaches were used for the GAMs fitting process: (1) using all presences and all theoretical absences pooled together, and (2) running ten iterations with all presences and 1000 theoretical absences selected randomly on each run, mimicking the “absences zones” approach used by Derville et al. 2016. We used a logit link for both approaches and allowed a maximum of three degrees of freedom for each spline to limit over-fitting (Becker et al., 2014). Furthermore, the model's effort was included as an offset term to account for variations in segment length. We fitted two different MAXENT approaches: (1) a classical approach with automatic features and (2) a more simplistic approach with only linear and quadratic features. We decided to test the second approach due to the virtual species simplistic response curves designed for the present study. We applied a regularization factor of 2 to avoid over-fitting.</li> </ul>
Model workflow	<p>We used 8-day grouped datasets to calibrate the models using a variety of modelling techniques. We applied a series of techniques to evaluate the model performance both in the geographical space and environmental space. We calculated the a test AUC with a random 5-fold approach (approx. 25% of dataset for testing) and the Schoener's D (Schoener, 1968) and a fuzzy similarity index based on Baroni-Urbani &amp; Buser's (1976) in the geographical space to compare model predictions with virtual species known predictions. Moreover, a niche overlap was tested in environmental space. The Schoener's D index and the Spearman correlation rank were applied to test for differences between modelling scenarios. All the process was repeated for 100 iterations.</p>
Software, codes and data	<ul style="list-style-type: none"> <li>• <b>Specify modelling platform:</b> dismo and mgcv R packages. Maxent 3.4.3</li> <li>• <b>Specify availability of codes:</b> N/A</li> <li>• <b>Specify availability of data:</b> N/A</li> </ul>
<b>DATA</b>	
Biodiversity data	<p>Virtual species, definitions can be found on the main paper and suppl. material.</p> <ul style="list-style-type: none"> <li>• <b>Details on taxonomic reference system:</b> N/A</li> <li>• <b>Ecological level:</b> population</li> <li>• <b>Biodiversity data source:</b> N/A</li> <li>• <b>Sampling design:</b> Surveys were performed around the islands and other vital areas, covering 39,316.9 km<sup>2</sup> per year. We used the transects designed by Faustino et al. (2010) to last four months (12 weeks) per year, with four days of sampling per week over four years (2013-2016). Cruise speed was defined as 7.5 knots during 8 hours each sampled day to avoid any area's resampling.</li> <li>• <b>Sample size per taxon:</b> Sample sized varied depending on the encounter rates applied. 5 different encounter rates per 100km were applied (1.4,1,0.8,0.6,0.2), referring to the number of groups of animals encountered per 100km sampled.</li> </ul>

	<ul style="list-style-type: none"> <li>• <b>Details on absence data collection:</b> we assumed the no detection of a species during the surveys as an absence.</li> </ul>
Data partitioning	<ul style="list-style-type: none"> <li>• <b>Selection of training data:</b> for the AUC validation we used 75% of data of training data.</li> <li>• <b>Selection of validation data:</b> for the AUC a 25%. Other methods rely on the comparison with the already known distribution of the virtual species.</li> </ul>
Predictor variables	<ul style="list-style-type: none"> <li>• <b>Predictor variables:</b> Depth and slope. Sea Surface Temperature (SST).</li> <li>• <b>Details on data sources:</b> Depth: derived from the digital elevation model of the EMODnet Bathymetry portal. Slope: calculated from the digital elevation model using QGIS 3.2. SST: obtained from the NASA's Multi-scale Ultra-high Resolution (MUR) downloaded on 01/12/2018 from <a href="https://mur.jpl.nasa.gov/">https://mur.jpl.nasa.gov/</a></li> <li>• <b>Spatial resolution and spatial extent:</b> all data were interpolated to a 2km grid resolution.</li> <li>• <b>Map projection:</b> WGS84 latlon</li> <li>• <b>Temporal resolution:</b> 8-day means for sea surface temperature</li> <li>• <b>Data processing:</b> Depth was extracted from the EMODnet bathymetry and up-scaled to a 2km resolution. Slope was calculated from the same source using QGIS 3.2. SST products were downloaded at a daily 1km resolution, we calculated 8-day estimates using the means of SST daily values. Finally, all the layers were aggregated to 2x2 km resolutions using a bilinear interpolation.</li> </ul>
<b>MODEL</b>	
Multicollinearity	We applied a variance inflation factor (VIF) approach as implemented in the R package usdm (Naimi, 2015) to test for collinearity. No warning signs were found.
Model settings	We used species presence/absence as the dependent variable, and so the logistic link function and binomial error term were used in the GAMs and GLMs. The effort was logarithmically transformed and used as an offset term. GLM response functions were modelled as a quadratic function. Two different MAXENT approaches were used: (1) a classical approach with automatic features and (2) a more simplistic approach with only linear and quadratic features. To avoid overfitting problems, we used a standard regularization parameter of 2.
Model estimates	<p><b>Assessment of model coefficients:</b> N/A</p> <p><b>Details on quantification of uncertainty in model coefficients:</b> N/A</p> <p><b>Assessment of variable importance:</b> N/A</p>
<b>ASSESSMENT</b>	

Performance statistics	<ul style="list-style-type: none"> <li>• <b>Performance statistics estimated on validation data:</b> AUC with 25% for testing purposes.</li> <li>• <b>Performance statistics estimated on test:</b> comparison with "known" virtual species projections in the geographic and environmental space, using the ENM-Tools.</li> </ul>
Plausibility check	<p>Response plots were calculated and compared with the original response curves for the virtual species.</p> <ul style="list-style-type: none"> <li>• <b>Expert judgements:</b> geographic projections were compared with the theoretical projected suitability. Visual comparison of predictions and theoretical suitability was done to better understand the performance of the models.</li> </ul>
<b>PREDICTION</b>	
Prediction output	Predictions of relative probability of presence expressed on a continuous scale.
Uncertainty quantification	N/A

1029  
1030