



HAL
open science

Identifying fish spawning grounds by combining catch declarations and scientific survey data

Baptiste Alglave, Youen Vermard, Etienne Rivot, Marie-Pierre Etienne,
Mathieu Woillez

► **To cite this version:**

Baptiste Alglave, Youen Vermard, Etienne Rivot, Marie-Pierre Etienne, Mathieu Woillez. Identifying fish spawning grounds by combining catch declarations and scientific survey data. 2023. hal-03674691v2

HAL Id: hal-03674691

<https://hal.science/hal-03674691v2>

Preprint submitted on 9 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Canadian Journal of Fisheries and Aquatic Sciences

Identifying mature fish aggregation areas during spawning season by combining catch declarations and scientific survey data

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2022-0110.R2
Manuscript Type:	Article
Date Submitted by the Author:	23-Dec-2022
Complete List of Authors:	Alglave, Baptiste; Ifremer Centre Atlantique, UMR DECOD; AGROCAMPUS OUEST, UMR DECOD Vermard, Youen; IFREMER, Unité EMH Rivot, Etienne; Agrocampus Ouest, UMR 0985 INRA / Agrocampus Ouest ESE Etienne, Marie-Pierre; IRMAR Woillez, Mathieu; Ifremer, STH/LBH
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)
Keyword:	Species distribution model, Spatio-temporal model, Hierarchical model, VMS and logbook data, Fish reproduction areas

SCHOLARONE™
Manuscripts

Identifying mature fish aggregation areas during spawning season by combining catch declarations and scientific survey data

Baptiste Alglave^{1,2*}, Youen Vermard¹, Etienne Rivot², Marie-Pierre Etienne³, & Mathieu Woillez⁴

¹ DECOD (Ecosystem Dynamics and Sustainability), IFREMER, Institut Agro, INRAE, Nantes, France

² DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, IFREMER, INRAE, Rennes, France

³ Mathematical Research Institute of Rennes IRMAR, Rennes University, Rennes, France

⁴ DECOD (Ecosystem Dynamics and Sustainability), IFREMER, Institut Agro, INRAE, Brest, France

*corresponding author: email: baptiste.alglave@agrocampus-ouest.fr; present address: DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, IFREMER, INRAE, Rennes, France

Abstract

Identifying and protecting essential fish habitats like spawning grounds requires an accurate knowledge of fish spatio-temporal distribution. Commercial declarations coupled with Vessel Monitoring System provide fine scale information on the full year to map fish distribution and identify essential habitats. We developed an integrated framework to infer fish spatial distribution on a monthly time step by combining scientific and commercial data while explicitly considering the preferential sampling of fisher towards areas of higher biomass. We developed a method to identify areas of persistent aggregation of biomass during the spawning season and interpret these as spawning areas. The model is applied to infer maps of relative biomass for three species (sole, whiting, squids) in the Bay of Biscay on a monthly time step over a 9-year period. Integrating several fleets in inference provides a good coverage of the area and improves model predictions. The preferential sampling parameters give insights into the temporal dynamics of the targeting behavior of the different fleets. Last, persistent aggregation areas reveal consistent with the available literature on spawning grounds, highlighting the potential of our approach to identify reproduction areas.

Keywords: Species distribution model, Spatio-temporal model, Hierarchical model, VMS and logbook data, Aggregation areas, Fish reproduction areas, Template Model Builder (TMB)

29 **1 INTRODUCTION**

30 Integrating fisheries into Marine Spatial Planning (MSP) to preserve ecosystem functions
31 and ensure a sustainable exploitation requires an accurate knowledge of fish spatio-
32 temporal distribution and more specifically of fish essential habitats such as reproduction
33 and nursery grounds (Janßen et al. 2018). However, such knowledge is still missing for
34 many species due to a lack of data with sufficient spatial, temporal or demographic
35 resolution (Delage and Le Pape 2016; Regimbart et al. 2018).

36 The available data to map fish distribution and identify essential habitats mainly rely on
37 either scientific survey data (fishery-independent data) or commercial data available
38 through on-board observer programs (fishery-dependent data) (Pennino et al. 2016). Both
39 data sources benefit of direct on-board recording of catches and are usually considered
40 as high quality data. Furthermore, both data sources were proved to be complementary
41 (Rufener et al. 2021). Scientific data benefit from a standardized sampling plan, a
42 standardized catchability and occur each year at the same period. Consequently, they
43 provide standardized data on a large spatial extent for most species and size classes
44 (Hilborn and Walters 2013; Nielsen 2015). Observer data potentially provide data over the
45 full year for all caught species, even though they do not follow a standardized protocol as
46 survey data. However, both scientific survey and onboard observer data are characterized
47 by a relatively low sampling intensity in space and time. Because of material limitations,
48 surveys occur only once or twice a year and provide a limited number of samples each
49 time (ICES 2005) and observer programs only cover a limited fraction of the entire fleet
50 (e.g. only 1% of all sea trips are covered by the French observer programs - Cornou et
51 al., 2021). The low sampling density of both data sources may lead to imprecise

52 predictions (ICES, 2005; Alglave et al., 2022) and constrains to consider only rough
53 temporal resolution (e.g. semesters, quarters or seasons – see for instance Kai et al.,
54 2017; Pinto et al., 2019; Rufener et al., 2021) to ensure a satisfying spatial coverage of
55 the data at each time step. However, the temporality of key biological events, such as the
56 reproduction peak, may be much tighter than the temporal resolution of data (Biggs et al.
57 2021). Hence, those data alone are likely not sufficient to provide accurate inferences on
58 essential fish habitats such as spawning grounds.

59 Commercial catch declarations combined with their fishing locations available from VMS
60 (Vessel Monitoring System) were proven to be an interesting alternative to obtain landing
61 per unit effort (LPUE) data with fine spatial and temporal resolution (Pedersen et al. 2009;
62 Bastardie et al. 2010; Gerritsen and Lordan 2010; Hintzen et al. 2012; Murray et al. 2013;
63 Azevedo and Silva 2020). However, considering commercial fisheries data to infer fish
64 spatial distribution remains highly challenging. Among other challenges, this implies
65 accounting for fisher sampling behavior. Fisher typically tend to preferentially sample
66 areas of higher biomass (a process referred to as preferential sampling, PS - Diggle et al.
67 2010). Hence, because data preferentially represent areas of highest biomass, ignoring
68 PS in the distribution of fishing effort when estimating spatial distribution on larger areas
69 can lead to overestimated biomass predictions (Conn et al. 2017; Pennino et al. 2019;
70 Alglave et al. 2022).

71 In a recent paper, Alglave et al. (2022) developed an integrated modeling framework to
72 infer spatial distribution of fish abundance by combining scientific survey CPUE and
73 commercial LPUE data while accounting for PS in the distribution of fishing effort. They
74 applied their framework to commercial data of a single month to match with the scientific
75 survey and did not consider any temporal dimension.

76 In this paper, we extend the modeling framework from Alglave et al. (2022) by adding a
77 temporal dimension to estimate fish spatio-temporal distribution at a monthly time step.
78 Our new model accounts for the variation over time (monthly time step) in the biomass
79 field as well as in the intensity of PS for distinct fishing fleets. To demonstrate the value of
80 the method, we selected and applied the model to 3 demersal species in the Bay of Biscay
81 (common sole, whiting and squids) characterized by contrasted configurations regarding
82 the available knowledge of their spawning grounds. We used those applications to
83 reinforce results obtained in Alglave et al. (2022) demonstrating how the integrated
84 framework allows to combine the information from several fleets in order to produce
85 accurate maps of spatio-temporal biomass. To illustrate the capacity of the framework to
86 identify areas of aggregation during the spawning season, we processed model outputs
87 to identify areas of recurrent aggregation occurring during the reproduction season and
88 compared these to the information available in the literature.

89

90 **2 MATERIAL AND METHODS**

91 In this section, we first present the different species, the datasets and how we process
92 and combine them to produce LPUE data in space and time. Second, we extend the model
93 proposed by Alglave et al. (2022) to introduce a temporal dimension on a discrete monthly
94 time step. In our applications, the models were fitted to data from 2010 to 2018 on a
95 monthly time step (108 time steps). Then, we illustrate how the PS component modifies
96 model predictions and can be interpreted, and how integrating several fleets in the
97 analysis further improves model predictions. Last, we detail the method used to
98 investigate spatio-temporal dynamics from model outputs and identify reproduction
99 grounds based on the aggregation patterns of each of three species.

100 **2.1 Case studies**

101 Sole is a data-rich case. Direct information about reproduction grounds is available
102 through egg and larvae surveys (Arbault et al. 1986). Reproduction period fall between
103 January and April, but the peak of the reproduction fall in February (Figure 7). Discard
104 rate is also very low, which makes the landings data a good proxy of the catch (ICES
105 2019a).

106 By contrast, Whiting is a data-poor case study where only indirect information of
107 reproduction period exists through spring trawl surveys (House and Forest 1993).
108 Reproduction period fall between March and May. Discard rates can be high (about 30%)
109 and thus landing data may provide a biased picture of the real catches (ICES 2019b).

110 Our third case study is Squids that represent a mixture of several species declared under
111 a common denomination in the catch (Loliginidae here referred as squids): *Loligo Vulgaris*
112 (Lamarck, 1798), *Loligo forbesii* (Steenstrup, 1856) and *Alloteuthis sp* (Lamarck, 1798).

113 Overall scientific survey suggests that the predominant species in the Bay of Biscay is
114 *Loligo Vulgaris* (ICES, 2020a, p.17). All 3 species are data-poor: no information exists
115 regarding their reproduction grounds but some information of the reproduction period
116 exists for *Loligo Vulgaris* (Moreno et al. 2002). For this species, the reproduction period
117 fall between January and April.

118 **2.2 Data**

119 **2.2.1 High spatial resolution catch per unit effort data for the mature component** 120 **of the populations**

121 We pre-processed the VMS and catch declaration (logbook) data to obtain high spatial
122 resolution LPUE data for the mature component of those three stocks, for three different
123 fishing fleet, and for each month of the 2010-2018 time series. In the text, we used the
124 term Landing Per Unit of Effort (LPUE) to refer to commercial observations expressed in
125 kg of (mature) biomass per hour fished. Discards are neglected in our approach, hence
126 LPUE are considered as biomass indices.

127 Our model can integrate data of different fishing fleets. For the purpose of our application,
128 we selected three different métiers that belong to the same fleet of trawlers (in this case,
129 the 'métier' term refers to a combination of gear and of a set of species that are targeted
130 by the vessels – see the Data Collection Framework and EC (2008)): OTB_DEF (bottom
131 otter trawl targeting demersal species), OTB_CEP (bottom otter trawl targeting
132 cephalopods) and OTT_DEF (multi-rig otter trawl targeting demersal species). In the
133 following those three métiers are referred as three fleets. These three fleets were selected
134 as they offer three main advantages. First, their targeting behaviors and technical
135 characteristics are similar. Second, catch per unit effort of trawlers are generally good
136 indicator of fish relative abundance while other gears (longline, gillnet) may face saturation

137 effects leading to non-linear relationship between catches and fishing time (Hovgêrd and
138 Lassen 2008). Third, the combination of the three fleets cover the full spatial domain
139 (Figure 1).

140 Because one of our primary goal is to identify spawning grounds, we filtered only the
141 mature fraction of the landings (i.e. the fraction of the individuals that can potentially
142 reproduce, not per se the fraction of the population that are spawning – this is detailed
143 further in the discussion). This was done by crossing the landings data with length class
144 and maturity data. For most of the landings, information on the commercial size categories
145 is available from the sales notes. These commercial categories are regularly sampled to
146 derive length structure of each commercial category. This allows us to estimate the
147 proportion of potentially mature fish in each commercial category by applying maturity
148 ogives and in turn estimate the proportion of mature fish for each landing declaration. See
149 SM1 for more detail. Note that this procedure was not possible for squids, as there are no
150 data on maturity and size classes for this species group.

151 Landing data were then combined with VMS data to finally obtained high spatial resolution
152 LPUE data discretized on a $0.05^\circ \times 0.05^\circ$ grid (i.e. 5.5 km x 3 km) on a monthly time step
153 (see SM2). This combination requires:

154 (1) to identify the fishing locations within the VMS data. This is realized individually
155 for each fishing vessel trajectory based on a speed threshold similarly as in
156 common data processing methods (Hintzen et al. 2012).

157 (2) to reallocate the logbooks declaration on the related VMS fishing locations. This
158 reallocation is realized individually for each fishing vessel trajectory by uniformly
159 reallocating the landings on all fishing locations. The link between both data

160 sources is realized through the combination 'vessel identifier x statistical
161 rectangle x fishing trip x day'. LPUE are then computed by simply dividing the
162 reallocated landings by the related fishing time.

163 **2.2.2 Scientific data**

164 We also integrated scientific data in the analysis. For whiting and squids we used the
165 survey data from the EVHOE survey. The Orhago survey was used for sole (ICES, 2020
166 - see SM, Figure S3). The data were extracted from the DATRAS database on the period
167 2010 - 2018. Only the mature fraction of the survey catches were kept in the analysis to
168 make it comparable with commercial data.

169 Orhago is an annual beam trawl survey occurring in November and designed to assess
170 sole stock status in the Bay of Biscay. Each year 50 stations are sampled within 4 strata
171 all along the Bay of Biscay. Note that this survey is mainly coastal and does not sample
172 offshore areas. EVHOE is an annual bottom trawl survey occurring in late October,
173 November and early December with a stratified sampling plan. It is designed for demersal
174 fishes in the Bay of Biscay and in the Celtic Sea. In the Bay of Biscay, 80 to 90 sampling
175 hauls are recorded each year.

176 **2.3 Spatio-temporal integrated model**

177 Alglave et al. (2022) developed a hierarchical integrated statistical model to infer spatial
178 distribution of fish density through scientific survey data and commercial data. It is
179 structured in 4 layers:

- 180 - the latent field that represents biomass spatial distribution, and that is the
181 main target of the inferences;

- 182 - the observations from scientific surveys and commercial declarations that
 183 are considered as direct zero-inflated observations of the latent field at the
 184 registered fishing locations;
- 185 - the fishing sampling intensity that relates fishing locations to the latent field
 186 and model explicitly the PS of commercial fleets towards areas of higher
 187 biomass;
- 188 - unknown parameters that control the shape of the biomass latent field and
 189 the sampling process.

190 This first model was purely spatial as no temporal dimension was included in the model.

191 In this paper, we extend the framework by incorporating a temporal component to model
 192 the evolution of the latent field of biomass across the monthly time steps (Figure 2).

193 **2.3.1 Biomass field**

194 As a notable extension of Alglave et al. (2022), the biomass field (eq.1) is modeled as a
 195 spatio-temporal Gaussian Random Field (GRF) through a log link as:

$$196 \quad \log(S(x,t)) = \alpha_S(t) + \delta(x,t) \quad (1)$$

197 where $x \in D \subset R^2$ stands for the spatial locations and $t \in \llbracket 1, T \rrbracket$ for the monthly time steps.

198 $S(x,t)$ is in the same unit as the data (here kg/hr fished as data are CPUE for survey trawls
 199 or LPUE for commercial landing declarations). The term $\alpha_S(t)$ is a time varying intercept

200 modeled as a fixed effect and $\delta(x,t)$ is a GRF spatio-temporal process which represents

201 the spatio-temporal correlation structure of the biomass field. As commercial data may not

202 always cover all time steps, the temporal correlation is a critical component that allows to

203 interpolate between time-steps. Here, the spatio-temporal term has a classical stationary

204 first-order autoregressive form (eq.2) following (Cameletti et al. 2013):

205
$$\delta(x,t) = \varphi \cdot \delta(x,t - 1) + \omega(x,t) \quad \text{for } t = 2, \dots, T \quad (2)$$

206 The autocorrelation coefficient φ is a scalar with $\varphi \in -1, 1$, $\omega(x,t)$ are spatial random
207 effects that represents the spatial innovation, modeled as a 0 mean GRF (with no temporal
208 correlation). Spatial random effects $\omega(x,t)$ are parameterized through a range parameter
209 ρ that corresponds to the distance at which spatial autocorrelation falls below 0.1.

210 Note that no covariate is included in the latent field to keep the model as simple as possible.
211 If any, the covariates effects are captured through the spatio-temporal term $\delta(x,t)$.
212 Similarly, the intercept $\alpha_S(t)$ was modeled through a simple fixed effect but more complex
213 specifications including some seasonal, yearly and interaction effects could be adopted
214 such as in Thorson et al. (2020).

215 **2.3.2 Sampling process for the commercial fishing points**

216 For most scientific surveys targeting a variety of species or covering wide areas, even if
217 the sampling plan is a stratified-random sampling and sampling takes into account the
218 most important commercial species in the area, the final distribution of the sampling
219 location can be considered independent from the biomass field distribution (see S3 for the
220 scientific survey used in this paper), and scientific sampling locations do not need to be
221 modeled explicitly (Diggle et al. 2010). By contrasts, the dependence between the
222 commercial fishing locations and the biomass field has to be modeled and included in the
223 likelihood function to capture preferential sampling. We extended the model proposed by
224 Alglave et al. (2022) to account for temporal variations in PS. Because the fishing behavior
225 is potentially different among fishing fleet, PS is modeled specifically for each fleet
226 denoted j (in the next section, j is also used to denote the scientific survey). Observed
227 fishing locations of any fishing fleet j are integrated in the likelihood through an

228 inhomogeneous point process (X_{comj} in the Figure 2) whose intensity $\lambda_j(x,t)$ (eq.3) controls
 229 the expected number of fishing points within a given area:

$$230 \quad \log(\lambda_j(x,t)) = \alpha_{X_j}(t) + b_j(t) \cdot \log(S(x,t)) + \eta_j(x,t) \quad (3)$$

231 where:

- 232 - the time varying intercept $\alpha_{X_j}(t)$ quantifies the average fishing intensity on the
 233 whole area; it is modeled as a fixed effect;
- 234 - the time varying $b_j(t)$ quantifies the strength of PS towards biomass $S(x,t)$; it is
 235 modeled as a fixed effect too. If $b_j(t) = 0$, then PS is null. If $b_j(t) > 0$, then PS
 236 occurs and the greater, the stronger PS. Alternatively, $b_j(t) < 0$ would indicate that
 237 fisher have a repulsive behavior towards the resource.
- 238 - the pure spatial GRF $\eta_j(x,t)$ captures the remaining spatial variability in the fishing
 239 point pattern not captured by the PS term (for instance, dependence of the fishing
 240 locations towards management regulations, distribution of other targeted species,
 241 habits/tradition).

242 **2.3.3 Observation process**

243 All observations for both scientific and commercial data of any fleet j are assumed all
 244 mutually independent conditionally on the latent field of biomass and the sampling
 245 locations. As data (both scientific and commercial) eventually present a high proportion of
 246 zero values, we model the observations through a Poisson-link zero-inflated model
 247 introduced by Thorson (2018) and already used in Alglave et al. (2022). The observation
 248 model explicitly considers that each fleet can have its own catchability and its own zero
 249 inflation parameter.

250 The probability to observe a catch data y_i conditionally on the location x_i (with i the
 251 observation index), the time-step t_i , the biomass field value $S(x_i, t_i)$ and the fleet j is
 252 expressed as follow:

$$253 \quad P(Y_i = y_i | x_i, t_i, S(x_i, t_i), j) = \begin{cases} p_i & \text{if } y_i = 0 \\ (1 - p_i) \cdot L\left(\frac{\mu_j(x_i, t_i)}{(1 - p_i)}, \sigma_j^2\right) & \text{if } y_i > 0 \end{cases} \quad (4)$$

$$254 \quad p_i = \exp(-e^{\xi_j} \cdot \mu_j(x_i, t_i)) \quad (5)$$

255 $\mu_j(x_i, t_i) = q_j \cdot S(x_i, t_i)$ is the expected catch of fleet j at location x_i and time step t_i . It is the
 256 product of the latent field value $S(x_i, t_i)$ and of the relative catchability coefficient specific
 257 for fleet j denoted q_j . Specifically for each fleet j , ξ_j is a zero-inflation parameter controlling
 258 the proportion of zero in the data, σ_j^2 is the observation variance when the catch is positive.
 259 Equation (4) shows the two components that compose the probability to observe a catch
 260 y_i :

- 261 • the probability to obtain a zero catch ($y_i = 0$). It is modeled as a Bernoulli variable
 262 with probability $p_i = \exp(-e^{\xi_j} \cdot \mu_j(x_i, t_i))$. p_i is equivalent to the probability to obtain a
 263 0 value with a Poisson distribution of intensity $e^{\xi_j} \cdot \mu_j(x_i, t_i)$. The value of ξ_j controls
 264 the intensity of the zero inflation, when ξ_j increases, the amount of zero in the data
 265 decreases. Then the probability to obtain a positive catch is given by $1 - p_i$.
- 266 • the value of the positive catch is modeled through a lognormal distribution L with
 267 expected value $\frac{\mu_j(x_i, t_i)}{(1 - p_i)}$ and observation error σ_j^2 . The standardization by $(1 - p_i)$
 268 allows to keep the expectancy of the observation model to $\mu_j(x)$.

269 The catchabilities q_j are not identifiable per se and some additional constraints need to be
270 set to estimate the relative catchability of each fleet (Alglave et al., (2022)). To ensure
271 identifiability, one fleet catchability is set as reference level (e.g. $q_{ref} = 1$, here OTB_DEF
272 was used as the reference fleet) and the other fleets' catchabilities are estimated relatively
273 to the reference fleet through the equation:

$$274 \quad q_j = k_j \cdot q_{ref} \quad (6)$$

275 **2.3.4 Maximum likelihood estimation**

276 The estimation of the spatio-temporal model is achieved through maximum likelihood
277 estimation. We used the Stochastic Partial Differential Equation (SPDE) approach that
278 allows to benefit from the nice computational properties of Gaussian Markov Random
279 Fields while working on a continuous domain (Lindgren et al. 2011). In practice our model
280 is coded with Template Model Builder (TMB - Kristensen et al., 2016) which benefits from
281 the Laplace approximation to integrate over random effects, automatic differentiation and
282 sparse matrix computation technics for a fast estimation of the model through maximum
283 likelihood estimation. Details on estimation are provided in SM 4, 5 and 6.

284 **2.4 Evaluating the interest of integrating multiple fleets**

285 Integrating several fleets in inference allows to cover the whole area (Figure 1) and is
286 expected to improve inferences. To illustrate the value of integrating the data from multiple
287 fleets within a single integrated model, we compared the spatial predictions obtained by
288 fitting the model to all available data with those obtained by integrating only one fleet. In
289 addition, we investigated if integrating all the fleets in inference increased the correlation
290 between scientific data and model predictions.

291 We also compared the coefficient of variation of the prediction between each model on a
292 single time step (here November 2018).

293 Note that scientific data was systematically integrated into inference (either in the
294 integrated model or in the single-fleet models). However, due to the low sample size
295 compared with intensive 'VMS x logbooks' data (about 80 scientific samples each year in
296 November compared with 17000 samples per month on average), they have very low
297 contribution to inference. This was extensively discussed in Alglave et al. (2022). Here
298 they mainly provide some standardized and reference data to assess the performance of
299 the framework.

300 **2.5 Evaluating the value of modeling PS**

301 **2.5.1 Comparing the inferences with and without PS**

302 We first assessed the impact of PS on the distribution of biomass by comparing
303 estimations obtained from integrated models (i.e. models fitted to all data sources)
304 accounting for PS with those obtained when ignoring PS. We computed the log-likelihood
305 related to each data source (commercial and scientific data) to assess if there is an
306 improvement in model goodness-of-fit when accounting or not for PS. Note that fitting a
307 model without PS is straightforward as it only requires to remove the sampling process
308 component from the likelihood function.

309 **2.5.2 Interpreting the intensity of preferential sampling**

310 The estimates of PS parameters $b_j(t)$ in eq. (3) may bring valuable information on the
311 dynamics of the fishery as they inform on the strength of the relationship between
312 commercial sampling distribution and species distribution. We investigate the variability of
313 the PS parameters among the three species and the different fleets. Then, focusing on

314 the sole case study, we highlight the insights brought by the model on the temporal
315 evolution of PS and its seasonal variations.

316 **2.6 Investigating spatio-temporal distribution and identifying reproduction** 317 **grounds**

318 The spatio-temporal model provides some insight on the temporal dynamics of species
319 distribution both at inter- and intra-annual levels. We applied a method to identify recurrent
320 aggregation areas from the maps of abundance inferred at each time step.

321 **2.6.1 Aggregation index**

322 We used the Getis and Ord index $G_d(x,t)$ (Getis and Ord 1992; Ord and Getis 1995) to
323 determine persistent aggregation areas (see for instance Milisenda et al., (2021)). The
324 generalized version of the Getis and Ord index is given in Bivand and Wong (2018) and
325 Ord and Getis (1995). Basically, $G_d(x,t)$ is a normalized version of the ratio between the
326 sum of the log-biomass (denoted $s(x,t)$) within a fixed neighborhood d and the sum of s
327 (x,t) on the entire area (for a fixed time step) (Getis and Ord 1992). We computed these
328 indices on $s(x,t) = \log(S(x,t))$ so that the $s(x,t)$ are Gaussian, which makes G_d Gaussian
329 too. In the application, we used a neighborhood distance $d=7.5$ km which defines a small
330 neighborhood of 8 cells (the direct neighbors of each cell grid) and allows to identify very
331 localized aggregation areas. Positive values for the aggregation index G_d indicates that s
332 (x,t) fall within a local patch of high values while negative G_d indicates that $s(x,t)$ fall within
333 a local patch of low values. Near 0 values G_d , indicates that $s(x,t)$ does not fall in some
334 local aggregation patch. As G_d follows a standardized Gaussian distribution, the
335 comparison between the value of the index and the quantiles of a standard Gaussian
336 distribution can be used to evaluate whether or not the latent field of biomass fall within a
337 statistically significant high or low aggregation patch. We used the quantile 99% (2.58) as

338 a threshold to ensure a high level of significance for patch detection (only local patch of
339 positive values are considered) and applied the Bonferroni correction to account for the
340 multiple statistical tests that are conducted.

341 Then, we define the persistence indices $IP(x,m)$ as the proportion of times point x falls
342 significantly within an aggregation area for a specific month/season m (can be either a
343 month or several months) among several years. Areas marked with high values of IP are
344 persistent aggregation areas throughout the time series.

345 **2.6.2 Comparing the results with the available literature**

346 Persistent aggregation areas derived from our model during the reproduction period (as
347 defined from the literature) were interpreted as potential recurrent reproduction grounds.
348 We compare those inferences with the information of reproduction ground available from
349 the literature (for sole and whiting).

350 Arbault et al. (1986) investigated the reproduction of sole along the Bay of Biscay based
351 on several egg surveys occurring in 1982. Five surveys were conducted between January
352 and May. Egg density was sampled in different locations from Hendaye to Pointe du Raz
353 ($43^{\circ}30'N$ - $48^{\circ}N$) and allowed to map the distribution of egg production on the full study
354 domain. The peak of reproduction occurred in February; thus we compare the maps
355 obtained from the February survey with the persistence index obtained from our model in
356 February.

357 For whiting, only two EVHOE trawl surveys occurred during spring (considered as the
358 reproduction period of whiting) between 1987 and 1992 in the Bay of Biscay (House and
359 Forest 1993). For each haul, the individuals were counted and aged. Individual of two
360 years and older were considered mature. We compare the distribution of mature

361 individuals obtained with these surveys and the index of persistence from our model
362 during spring (March to May).

363 No available information exists regarding the reproduction grounds of squids in this area,
364 however the study from Moreno et al. (2002) investigated the reproduction period for
365 *Loligo vulgaris* in the Eastern Atlantic and highlighted that their reproduction falls in winter
366 and spring with a peak from January to April. We compute the persistence index for this
367 period to identify the spatial aggregation patterns that emerge from the model outputs and
368 that could be considered as spawning grounds.

369 To assess whether the aggregation patterns within the reproduction period are stable over
370 the time period, we iteratively computed the persistence index over a 5-year mobile time-
371 span while pushing forward one year each time.

372 **3 RESULTS**

373 **3.1 Assessing the contribution of each data sources to inference**

374 Results highlight how combining several commercial fleets in the framework brings a
375 better picture of the spatial distribution on the whole domain. For instance, when
376 comparing model predictions with survey data (for the month of the survey), integrating
377 several fleets into the analysis improves correlation with scientific data (Figure 3). It also
378 reduces the standard deviation of the predictions on the full domain (Figure S7).

379 When looking at the predictions within the spatial range of the fleets, single-fleet models
380 logically provide similar spatial predictions compared with the integrated model (Figure 4;
381 red dots). However, when using single fleet data, predictions realized outside the spatial
382 range of the fleet largely depart from the ones realized through the integrated models
383 (black dots, Figure 4), emphasizing the contribution of the other fleets to improve
384 inferences on areas poorly covered by single fleet. This is particularly evidenced with the
385 OTB_CEP and OTT_DEF fleets that partially cover the study area compared with
386 OTB_DEF that better cover the whole study area (Figure 1).

387 **3.2 Interpreting estimates PS intensity**

388 Estimates of the PS intensity (b parameters in eq. (3)) for the different species, the
389 different fleets and the different time steps provide information on the targeting behavior
390 that are consistent with expertise. Estimates of b are positive for each species and each
391 fleet (Figure 5, left column). For squids, PS is the strongest for OTB_CEP followed by
392 OTB_DEF and OTT_DEF. This is consistent with the expert knowledge of the targeting
393 behavior of these fleets: OTB_CEP target cephalopods and catch on average 15% of
394 squids while OTB_DEF and OTT_DEF catch respectively 5% and 1% of squid). A similar
395 pattern can be identified for whiting ($b_{OTB_{CEP}} > b_{OTB} > b_{OTT}$); this is consistent with species

396 spatial distribution as whiting (like squids) are found in coastal areas (Figure 7) where the
397 OTB_CEP fleet is preferentially operating (Figure 1). For sole, the strength of PS is on
398 average higher for OTB_CEP and OTT_DEF than for OTB_DEF but with less contrast
399 between the three commercial fleets which is also consistent with expertise as those three
400 fleet target this high commercial value species.

401 Interestingly, some of the b parameters time series emphasize seasonal patterns (Figure
402 6, top). For instance, in the sole case study for the OTB_CEP fleet, the b parameters are
403 higher in summer and autumn emphasizing relatively stronger PS, while being lower in
404 winter and early spring (but see section 3.4 below for a more detailed interpretation of this
405 seasonality pattern).

406 **3.3 Evaluating the influence of PS on spatial inferences**

407 Because estimates of b are positive, spatial density of fishing points is positively correlated
408 with biomass density. Then logically, ignoring PS leads to a positive bias (overestimate)
409 in biomass estimates in areas not sampled by the commercial fleets compared to
410 estimates obtained while considering (Figure 5, right column, black points), but does not
411 strongly affect predictions in locations within the range of the fleets (blue points).
412 Considering PS only slightly improves the fit of the model to the data. For the Sole case
413 study, some improvement of the likelihood occurred for both the likelihood associated with
414 the commercial and the scientific data (Table 1). For whiting and squids, there are no
415 strong modifications in both scientific and commercial likelihoods.

416 **3.4 Investigating spatio-temporal dynamics of fish biomass**

417 Results provide biomass density maps on a monthly time step that emphasize seasonal
418 distribution patterns and from which aggregation index were calculated. The temporal
419 correlation parameter (φ) is estimated around 0.8 for all the species emphasizing strong

420 between months temporal correlations in the biomass field values. The range parameters
421 are estimated to 55 km for sole and squids while being estimated to 67 km for whiting
422 emphasizing wider spatial autocorrelation for this species.

423 Concerning the sole case study, model predictions highlight the relatively offshore
424 distribution from November to April and a more coastal distribution from June to October
425 suggesting some offshore-coastal migrations between these 2 periods (Figure 6, bottom).
426 In particular, the migration in June/July is associated with a contraction of the sole
427 distribution around the Vendée coast, the Gironde Estuary and the Landes coast (45.5°N-
428 46°N) while the migration in November leads to an expansion of the species distribution
429 towards the offshore areas all along the Bay of Biscay. Interestingly, such seasonality
430 coincides with the seasonality of PS intensity for the OTB_CEP (Top of Figure 6 and SM
431 9). Higher PS parameter values are associated with a coastal distribution of sole while
432 lower values correspond to offshore distribution of sole.

433 Similar maps can be computed for the other species and are presented in SM10.

434 **3.5 Aggregation index and reproduction grounds**

435 For both sole and squids, areas of aggregation persistent over the years identified during
436 the spawning period exhibit strong aggregation patterns that match the available
437 knowledge of their reproduction grounds. For sole, the aggregation areas globally match
438 with the observed area of maximum egg concentration (Figure 7), although the spawning
439 grounds identified by egg maps are slightly further East of those identified by our method.
440 This slight discrepancy could be interpreted as an effect of the larval drift as the maps
441 provided by Arbault et al. (1986) are concentration of eggs and not reproduction grounds
442 per se. This is consistent with the simulation analysis of Ramzi et al. (2001) showing that

443 the eggs and larval drift in this area of the Biscay Bay is oriented to the East. Overall,
444 these aggregation areas are stable over time (Figure 8).

445 For whiting, similar patterns can be identified during the reproduction period (Figure 7);
446 they match with previous studies investigating the spatial distribution of mature whittings
447 (House and Forest 1993). In particular, the Northern (3°W - 47°N) and the Southern (2°W -
448 45.5°N) aggregation patches are almost systematically significantly considered as
449 aggregation areas (aggregation index equals 1) while the other middle one (2.5°W - 46.5°N)
450 is classified as an aggregation area that appears less frequently over the years. An
451 additional persistent aggregation area can be identified in the North of the Bay of Biscay
452 (4.5°W - 48°N) suggesting that reproduction may also occur in this area which was not
453 identified in the report of House and Forest (1993). Interestingly, the Northern aggregation
454 area (3°W - 47°N) is more intense at the end of the period (Figure 8).

455 For squids, no information related to any reproduction ground exists in the literature, only
456 the time period of the reproduction is known (the peak fall between January to April). On
457 this time period, some persistent aggregation areas can be evidenced in coastal areas
458 (Figure 7) along the Vendée coast (2.5°W - 46.5°N), the Landes coasts (1.5°W - 44°N to
459 45°N) and around Belle-Île-en-Mer (3°W - 47.25°N). Interestingly, the two Northern
460 aggregation areas are more intense at the end of the time series compared to the
461 beginning of the time series (Figure 8).

462 Maps of persistent aggregation areas are available for all month and evidence some other
463 aggregation areas outside of the reproduction period (SM11). For instance, for sole a
464 persistent patch can be identified offshore the Gironde Estuary (1.5°W – 45.5°N) from
465 August to December.

466 **4 DISCUSSION**

467 **Main findings**

468 In this paper, we develop a framework to infer fish spatio-temporal distribution on a
469 monthly time step while combining scientific survey data and commercial catch
470 declarations from several fleets. Commercial catch data constitute a valuable data source
471 that complements scientific survey or onboard sampling programs by providing much
472 higher spatio-temporal sampling density. Those complementary sources of data were
473 integrated through a spatio-temporal hierarchical model taking into account spatio-
474 temporal variation within the biomass field and PS on a monthly time step. We fitted the
475 model to VMS-logbooks data filtered and processed over the period 2010-2018 for 3
476 demersal species (sole, squids and whiting) in the Bay of Biscay.

477 We emphasize the benefit of integrating several spatially complementary fleets to infer
478 fish distribution throughout the year. We demonstrate how the within-year dynamic of the
479 PS parameters can be interpreted with regards to the joint dynamics of species distribution
480 and fishing distribution and to the overall targeting behavior of the fleets (e.g. OTB_CEP
481 for the squids case study). Even though PS parameters are not fishing intention per se
482 (Bourdaud et al. 2019), these could advantageously complement information provided by
483 landing profiles to estimate the targeting behavior of any group of vessels (either
484 métier/fleet or any group that would seem appropriate).

485 Interestingly, although interpretation of the PS parameters provide insight into the spatio-
486 temporal fleet dynamics, accounting for PS in the inferences does not significantly improve
487 model fitting even when some fleets emphasize strong PS (e.g. squids, OTB_CEP). These
488 results contrast with Alglave et al. (2022), and could result from the integration of several

489 fleets in the analysis that allow a full coverage of the area. Indeed, in Alglave et al. (2022),
490 the fleet emphasizing strong PS only covered a restricted (and coastal) part of the area.
491 As introducing PS mainly affects inferences on poorly sampled areas, predictions in the
492 offshore areas where mostly affected. Here, as the fleets are all estimated to have a
493 positive PS and cover the whole area, PS only downscale the predictions in the few areas
494 that remain unsampled.

495 Filtering the mature fraction of the population in both the scientific and the commercial
496 data make possible to infer the spatio-temporal distribution of the fraction of the biomass
497 that could be potentially mature through the year on a monthly time step. We developed
498 an index to infer aggregation areas of the potentially mature fraction of the biomass that
499 are persistent across years. When calculated on a temporal window predefined following
500 the available information on the reproduction period for each species, the aggregation
501 index enables to identify the main recurrent spatial aggregation areas within the
502 reproduction period. Results demonstrate that the recurrent aggregation areas identified
503 from our method for Sole and Whiting were highly consistent with those already identified
504 in the literature. Our results demonstrate how the aggregation index can provide new
505 insights on the spawning grounds for species like squids for which no information on the
506 spawning grounds is available on the literature. Areas of high aggregation persistent
507 across years were identified during the expected period of reproduction and could be
508 interpreted as spawning grounds. This opens perspectives for applying more
509 systematically the approach for species where no information of reproduction grounds is
510 available to fill the gaps in our knowledge with minimum cost (Delage and Le Pape 2016;
511 Regimbart et al. 2018).

512

513

514 **Combining our results with other data sources to refine the identification of**
515 **spawning grounds**

516 Persistent aggregation areas should be considered as potential spawning areas rather
517 than actual spawning areas. Indeed, although the mature fraction of the biomass was
518 filtered in the data, our maps do not directly inform whether individuals are actually
519 reproducing or not. The outputs of the model provide maps of the mature fraction of the
520 population (i.e. the individuals that can reproduce) and not the spawning fraction of the
521 population (i.e. the individuals that are actually spawning). However, by focusing on the
522 temporal window identified as reproduction period in the literature, we limit the risk of
523 misinterpreting the aggregation areas as reproduction areas.

524 Our results can also be used to help gathering additional data to identify reproduction
525 grounds. Typically, our maps could be of great help to design surveys to record eggs,
526 larvae and spawning individuals that would provide direct information of species
527 reproduction (Fox et al. 2008). Because developing such additional surveys would be
528 highly expensive, our maps could provide valuable a priori information to optimize the
529 survey design and potentially find a compromise between the cost, the spatial extent, the
530 temporal coverage of the survey and the accuracy of the expected estimates/predictions.
531 Similar ideas were already applied to the sole case study to investigate more precisely
532 the space-time variation of sole reproduction. Arbault et al. (1986) work provided a priori
533 information of reproduction grounds that allowed designing more localized surveys to
534 study inter- and intra-annual variability of one specific sole spawning area (Petitgas 1997).
535 Several statistical methods have been developed since and are suitable to optimize such
536 adaptive sampling design; see for instance the recent work of Leach et al. (2021).

537 Our results could also be combined with fisher expert knowledge (Yochum et al. 2011) to
538 complement our knowledge of fish reproduction (Delage and Le Pape 2016). For instance,
539 Bezerra et al. (2021) and Silvano et al. (2006) proved the usefulness of fishers knowledge
540 to determine the temporality of fish spawning and to identify some spawning grounds by
541 crossing the information of aggregation areas provided by several fisher. These were
542 proved complementary with scientific data as they can be available at low cost and provide
543 local knowledge of fish ecology.

544 **Limits and perspectives for the approach**

545 Our framework has several limitations that are all material for future research avenues.
546 Our model remains relatively simple with regards to all the temporal processes that
547 actually occur within a fishery. It is both a strength and a weakness: the model remains
548 relatively generic, but one might want to extend it further to account for other temporal and
549 spatio-temporal processes affecting fisheries dynamics. For instance, we opted for a non-
550 seasonal representation of the model. One could make it seasonal by decomposing the
551 intercepts $\alpha_S(t)$ and $\alpha_{X_j}(t)$ as well as the random effects $\delta(x,t)$ and $\eta(x,t)$ into yearly and
552 seasonal terms in addition to some 'season x year' interaction terms as performed in
553 Thorson et al. (2020). In their work, such specification mainly allowed to provide
554 information over the time-steps where data was lacking. In the configuration of our case
555 studies, data is available for all time steps and have a relatively good coverage of the
556 study domain. Consequently, even though it provides a nice conceptual view of
557 seasonality, complexifying our model in that direction should not deeply modify our
558 inference of the biomass field. Alternatively, our framework could integrate orthogonal
559 spatio-temporal terms in the latent field to capture the main mode of variability of the

560 biomass field (Thorson et al. 2020b). Such orthogonal terms would allow to capture the
561 main spatial patterns that structure the latent field as well as their variation in time. These
562 could prove very useful to identify the structuring processes that affect species distribution
563 and could give a valuable insight in the space-time dynamics of the species. Another
564 exciting research avenue would consist in integrating population dynamics in the latent
565 field of biomass (Cao et al. 2020). This would require to refine further the demographic
566 resolution of the VMS-logbooks data (see for instance Azevedo and Silva 2020), but once
567 done, it would give access to huge data for inferring the space-time dynamics of fish
568 populations. Finally, our model considers fisher preferentially sample areas where the
569 biomass is higher (preferential sampling), but does not consider any other drivers and
570 specifically the temporal and spatio-temporal relations that can affect fishers behavior.
571 These can be highly complex and may depend on the distribution of the resource,
572 tradition/habits, management regulations (Abbott et al., 2015; Girardin et al., 2017; Salas
573 and Gaertner, 2004; Hintzen, 2021). These drivers are rarely studied in both space and
574 time (although see Tidd et al., 2015). Our framework could allow to jointly model the
575 dynamics of the species, the distribution of the effort, the link that relates species
576 distribution and effort in space/time and all the other spatial and/or temporal drivers that
577 affect the distribution of fishing effort. For instance, we could relate the fishing intensity to
578 the biomass field from the previous time steps, or alternatively consider that the locational
579 choice depends on the catches of the previous time steps. Adding such covariates and
580 spatio-temporal dependencies in the sampling equation (eq.3) will probably not modify the
581 overall pattern of biomass distribution, but it would make possible to quantify the drivers
582 of fisher behavior and give valuable insight to the fishery dynamics.

583 Including discards would potentially improve our approach. Indeed, logbooks data are
584 landings declarations data which means they inform on the landings and not on the true
585 catch. Thus, by assuming the landings per unit effort are proportional to the biomass, we
586 make the hypothesis that the discard rate is constant in space and time and does not
587 affect model predictions. This should not be a problem for sole and squids as the discards
588 are low and TAC have not been really binding during the studied period. However, the
589 issue might be more stringent for whiting and/or other species with a high and non-
590 stationary level of discards. Integrating discards data in the analysis could help solving
591 this issue. Stock et al. (2019) and Yan et al. (2022) used observer data to model bycatch
592 in both space and time and Breivik et al. (2017) used bycatch data from onboard surveys
593 to predict the temporal evolution of bycatch realized in the full commercial data. Similarly,
594 we could integrate into the same analysis the logbooks and the observer data by assuming
595 that the catch of observer represents the sum of landings (which is also observed in the
596 logbooks data) and discards (which is unobserved in the logbooks data). This way, the
597 discards information available from observer data would be shared with the logbooks data
598 and would allow correcting for the missing portion of catch declarations data while possibly
599 accounting for possible space or time variation in the discard rate.

600 Our analysis rely on the hypothesis that the spawning season is known a priori. Extending
601 the approach to infer the spawning season based on the temporal dynamics of the
602 aggregation patterns could improve our knowledge of species spatio-temporal distribution.
603 In particular, identifying the main species phenomenological phases and their consistency
604 (or shift) in time is crucial in the context of global change (Thorson et al. 2020a). In our
605 study, we computed the aggregation index on a predefined temporal window based on
606 literature assumed to be the reproduction period (Arbault et al. 1986; Houise and Forest

607 1993; Moreno et al. 2002). Several methods exist and could be adapted to extract the
608 spatial patterns that shape model outputs, their related temporal variation and identify the
609 main phenological phases that characterize species distribution (e.g. reproduction,
610 feeding - see for instance Empirical Orthogonal Functions or Principal Oscillation Patterns
611 - Cressie and Wikle 2015; Wikle et al. 2019). While the approach we adopted in the
612 manuscript requires to know the reproduction period of the species and would be
613 inappropriate in a context of a changing reproduction time-span, those alternative
614 methods would not require any a priori. Hence, these methods would be more appropriate
615 to identify phenological modifications in species life cycle in response to climate change.
616 Applying those kind of methods to the huge amount of data available from mandatory
617 declarations, would make possible to track the effect of climate change on fish phenology
618 at a monthly/seasonal scale, while it is generally only possible at a yearly time step
619 through scientific survey data (Maureaud et al. 2020).

620 Last, confidentiality remains a major limitation to the massive use of VMS data (Hintzen
621 2021). Indeed, there are strong confidentiality constraints on these data due to the huge
622 information available on fisher fishing grounds. Few countries are now giving free access
623 to their data (e.g. Norway), but in most cases administrative procedures to get access to
624 the data remain a burden and still constitute a limitation for the use of 'VMS x logbooks'
625 data for routine operational use.

626 **Future use for Marine Spatial Planning**

627 Our model has potential application for Marine Spatial Planning (MSP). Janßen et al.
628 (2018) highlighted that one of the main requirements for implementing MSP is the
629 availability of fine scale information on species distribution and of their essential habitats.
630 Here we propose a method which can provide such information for the fraction of the

631 population available through catch declarations (i.e. mainly the adult fraction and in some
632 cases part of the juvenile fraction). This knowledge is required to design Marine Protected
633 Areas (MPA – see for instance Lambert et al. (2017) or Loiseau et al. (2003)), Fishery
634 Conservation Zones (Delage et Le Pape, 2016 ; Regimbart et al., 2018), or alternatively
635 identify areas that should be kept for fishing in a context where many other human
636 activities are competing in space and time with fishing (Campbell et al. 2014; Bastardie et
637 al. 2015). This would require to integrate our results into bio-economic models in order to
638 evaluate alternative management regulations and assess their tradeoffs in regards to all
639 the sets of ecosystem services provided through activities such as fishing, aquaculture,
640 energy, shipping, recreation and conservation (Nielsen et al. 2018).

641

Draft

642 **ACKNOWLEDGMENT**

643 The authors acknowledge the Pôle de Calcul et de Données Marines (PCDM;
644 [https://wwz.ifremer.fr/en/Research-Technology/Research-Infrastructures/Digital-
645 infrastructures/Computation-Centre](https://wwz.ifremer.fr/en/Research-Technology/Research-Infrastructures/Digital-
645 infrastructures/Computation-Centre)) for providing DATARMOR supercomputer on which
646 the model has been fitted.

647 The authors are grateful to the Direction des pêches maritimes et de l'aquaculture (DPMA)
648 and Ifremer (Système d'Informations Halieutiques - SIH) who provided the aggregated
649 VMS and logbooks data. The findings and conclusions of the present paper are those of
650 the authors.

651 **FUNDINGS**

652 The authors declare no specific funding for this work.

653 **COMPETING INTERESTS**

654 The authors declare there are no competing interests.

655 **DATA AVAILABILITY STATEMENT**

656 Survey data are available through the DATRAS portal ([https://www.ices.dk/data/data-
657 portals/Pages/DATRAS.aspx](https://www.ices.dk/data/data-
657 portals/Pages/DATRAS.aspx)) with the package 'icesDatras' ([https://cran.r-
658 project.org/web/packages/icesDatras/index.html](https://cran.r-
658 project.org/web/packages/icesDatras/index.html)). Logbooks and VMS data are
659 confidential data and they are available on specific request to DPMA.

660 **CODES**

661 Toy example codes of the model are available on the github link:

662 https://github.com/balglave/sdm_vms_logbooks

663 **REFERENCES**

- 664 Abbott, J., Haynie, A., and Reimer, M. 2015. Hidden Flexibility: Institutions, Incentives, and
665 the Margins of Selectivity in Fishing. *Land Economics* **91**: 169–195.
666 doi:10.3368/le.91.1.169.
- 667 Alglave, B., Rivot, E., Etienne, M.-P., Woillez, M., Thorson, J.T., and Vermard, Y. 2022.
668 Combining scientific survey and commercial catch data to map fish distribution. *ICES*
669 *Journal of Marine Science: fsac032*. doi:10.1093/icesjms/fsac032.
- 670 Arbault, P.S., Camus, P., and Bec, C. le. 1986. Estimation du stock de sole (*Solea vulgaris*,
671 Quensel 1806) dans le Golfe de Gascogne à partir de la production d'œufs. *Journal of*
672 *Applied Ichthyology* **2**(4): 145–156. doi:10.1111/j.1439-0426.1986.tb00656.x.
- 673 Azevedo, M., and Silva, C. 2020. A framework to investigate fishery dynamics and species
674 size and age spatio-temporal distribution patterns based on daily resolution data: a
675 case study using Northeast Atlantic horse mackerel. *ICES Journal of Marine Science*
676 **77**(7–8): 2933–2944. doi:10.1093/icesjms/fsaa170.
- 677 Bastardie, F., Nielsen, J.R., Eigaard, O.R., Fock, H.O., Jonsson, P., and Bartolino, V. 2015.
678 Competition for marine space: modelling the Baltic Sea fisheries and effort
679 displacement under spatial restrictions. *ICES Journal of Marine Science* **72**(3): 824–
680 840. doi:10.1093/icesjms/fsu215.
- 681 Bastardie, F., Nielsen, J.R., Ulrich, C., Egekvist, J., and Degel, H. 2010. Detailed mapping of
682 fishing effort and landings by coupling fishing logbooks with satellite-recorded
683 vessel geo-location. *Fisheries Research* **106**(1): 41–53.
- 684 Bezerra, I.M., Hostim-Silva, M., Teixeira, J.L.S., Hackradt, C.W., Félix-Hackradt, F.C., and
685 Schiavetti, A. 2021. Spatial and temporal patterns of spawning aggregations of fish
686 from the Epinephelidae and Lutjanidae families: An analysis by the local ecological
687 knowledge of fishermen in the Tropical Southwestern Atlantic. *Fisheries Research*
688 **239**: 105937.
- 689 Biggs, C.R., Heyman, W.D., Farmer, N.A., Kobara, S., Bolser, D.G., Robinson, J., Lowerre-
690 Barbieri, S.K., and Erisman, B.E. 2021. The importance of spawning behavior in
691 understanding the vulnerability of exploited marine fishes in the US Gulf of Mexico.
692 *PeerJ* **9**: e11814.
- 693 Bivand, R.S., and Wong, D.W.S. 2018. Comparing implementations of global and local
694 indicators of spatial association. *TEST* **27**(3): 716–748. doi:10.1007/s11749-018-
695 0599-x.
- 696 Bourdaud, P., Travers-Trolet, M., Vermard, Y., and Marchal, P. 2019. Improving the
697 interpretation of fishing effort and pressures in mixed fisheries using spatial overlap
698 metrics. *Can. J. Fish. Aquat. Sci.* **76**(4): 586–596. doi:10.1139/cjfas-2017-0529.

- 699 Breivik, O.N., Storvik, G., and Nedreaas, K. 2017. Latent Gaussian models to predict historical
700 bycatch in commercial fishery. *Fisheries Research* **185**: 62–72.
- 701 Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. 2013. Spatio-temporal modeling of
702 particulate matter concentration through the SPDE approach. *AStA Adv Stat Anal*
703 **97**(2): 109–131. doi:10.1007/s10182-012-0196-3.
- 704 Campbell, M.S., Stehfest, K.M., Votier, S.C., and Hall-Spencer, J.M. 2014. Mapping fisheries for
705 marine spatial planning: Gear-specific vessel monitoring system (VMS), marine
706 conservation and offshore renewable energy. *Marine Policy* **45**: 293–300.
707 doi:10.1016/j.marpol.2013.09.015.
- 708 Cao, J., Thorson, J.T., Punt, A.E., and Szuwalski, C. 2020. A novel spatiotemporal stock
709 assessment framework to better address fine-scale species distributions:
710 development and simulation testing. *Fish and Fisheries* **21**(2): 350–367.
- 711 Conn, P.B., Thorson, J.T., and Johnson, D.S. 2017. Confronting preferential sampling when
712 analysing population distributions: diagnosis and model-based triage. *Methods in*
713 *Ecology and Evolution* **8**(11): 1535–1546.
- 714 Cornou, A.-S., Quinio-Scavinner, M., Sagan, J., Cloâtre, T., Dubroca, L., and Billet, N. 2021.
715 Captures et rejets des métiers de pêche français - Résultats des observations à bord
716 des navires de pêche professionnelle en 2019. Ifremer.
- 717 Cressie, N., and Wikle, C.K. 2015. *Statistics for spatio-temporal data*. John Wiley & Sons.
- 718 Delage, N., and Le Pape, O. 2016. Inventaire des zones fonctionnelles pour les ressources
719 halieutiques dans les eaux sous souveraineté française. Première partie: Définitions,
720 critères d'importance et méthode pour déterminer des zones d'importance à
721 protéger en priorité. Rapport de recherche, Pôle halieutique AGROCAMPUS OUEST,
722 Rennes.
- 723 Diggle, P.J., Menezes, R., and Su, T. 2010. Geostatistical inference under preferential
724 sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**(2):
725 191–232.
- 726 EC. 2008. 2008/949/EC: Commission Decision of 6 November 2008 adopting a multiannual
727 Community programme pursuant to Council Regulation (EC) No 199/2008
728 establishing a Community framework for the collection, management and use of data
729 in the fisheries sector and support for scientific advice regarding the common
730 fisheries policy. *In* OJ L. Available from
731 <http://data.europa.eu/eli/dec/2008/949/oj/eng> [accessed 23 December 2022].
- 732 Fox, C.J., Taylor, M., Dickey-Collas, M., Fossum, P., Kraus, G., Rohlf, N., Munk, P., van Damme,
733 C.J., Bolle, L.J., and Maxwell, D.L. 2008. Mapping the spawning grounds of North Sea
734 cod (*Gadus morhua*) by direct and indirect means. *Proceedings of the Royal Society*
735 *B: Biological Sciences* **275**(1642): 1543–1548.
- 736 Gerritsen, H., and Lordan, C. 2010. Integrating vessel monitoring systems (VMS) data with
737 daily catch data from logbooks to explore the spatial distribution of catch and effort
738 at high resolution. *ICES Journal of Marine Science* **68**(1): 245–252.
- 739 Getis, A., and Ord, J. 1992. The analysis of spatial association by use of distance statistics.
740 *Geographical Analysis*.
- 741 Girardin, R., Hamon, K.G., Pinnegar, J., Poos, J.J., Thébaud, O., Tidd, A., Vermard, Y., and
742 Marchal, P. 2017. Thirty years of fleet dynamics modelling using discrete-choice
743 models: What have we learned? *Fish and Fisheries* **18**(4): 638–655.
744 doi:https://doi.org/10.1111/faf.12194.

- 745 Hilborn, R., and Walters, C.J. 2013. Quantitative Fisheries Stock Assessment: Choice,
746 Dynamics and Uncertainty. Springer Science & Business Media.
- 747 Hintzen, N.T. 2021. Zooming into small-scale fishing patterns: The use of vessel monitoring
748 by satellite in fisheries science. PhD Thesis, Wageningen University.
- 749 Hintzen, N.T., Bastardie, F., Beare, D., Piet, G.J., Ulrich, C., Deporte, N., Egekvist, J., and Degel,
750 H. 2012. VMStools: Open-source software for the processing, analysis and
751 visualisation of fisheries logbook and VMS data. *Fisheries Research* **115**: 31–43.
752 Elsevier.
- 753 Houise, C., and Forest, A. 1993. Etude de la population du Merlan (*Merlangius merlangius*
754 L.) du Golfe de Gascogne. Ifremer.
- 755 Hovgêrd, H., and Lassen, H. 2008. Manual on estimation of selectivity for gillnet and longline
756 gears in abundance surveys. Food & Agriculture Org.
- 757 ICES. 2005. Report of the Workshop on Survey Design and Data Analysis (WKSAD). Sète,
758 France.
- 759 ICES. 2019a. Sole (*Solea solea*) in divisions 8.a–b (northern and central Bay of Biscay).
760 Advice. Available from <https://doi.org/10.17895/ices.advice.4775>.
- 761 ICES. 2019b. Whiting (*Merlangius merlangus*) in Subarea 8 and Division 9.a (Bay of Biscay
762 and Atlantic Iberian waters). doi:10.17895/ICES.ADVICE.4777.
- 763 ICES. 2020a. Working Group on Cephalopod Fisheries and Life History (WGCEPH). ICES.
764 doi:10.17895/ICES.PUB.6032.
- 765 ICES. 2020b. International Bottom Trawl Survey Working Group (IBTSWG). ICES Scientific
766 Reports, ICES. Available from <http://www.ices.dk/sites/pub/Publication>
767 Reports/Forms/DispForm.aspx?ID=37066 [accessed 28 May 2021].
- 768 Janßen, H., Bastardie, F., Eero, M., Hamon, K.G., Hinrichsen, H.-H., Marchal, P., Nielsen, J.R., Le
769 Pape, O., Schulze, T., and Simons, S. 2018. Integration of fisheries into marine spatial
770 planning: Quo vadis? *Estuarine, Coastal and Shelf Science* **201**: 105–113.
- 771 Kai, M., Thorson, J.T., Piner, K.R., and Maunder, M.N. 2017. Spatiotemporal variation in size-
772 structured populations using fishery data: an application to shortfin mako (*Isurus*
773 *oxyrinchus*) in the Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*
774 **74**(11): 1765–1780.
- 775 Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic
776 Differentiation and Laplace Approximation. *Journal of Statistical Software* **70**(1): 1–
777 21. doi:10.18637/jss.v070.i05.
- 778 Lambert, C., Virgili, A., Pettex, E., Delavenne, J., Toison, V., Blanck, A., and Ridoux, V. 2017.
779 Habitat modelling predictions highlight seasonal relevance of Marine Protected
780 Areas for marine megafauna. *Deep Sea Research Part II: Topical Studies in*
781 *Oceanography* **141**: 262–274.
- 782 Leach, C.B., Williams, P.J., Eisaguirre, J.M., Womble, J.N., Bower, M.R., and Hooten, M.B. 2021.
783 Recursive Bayesian computation facilitates adaptive optimal design in ecological
784 studies. *Ecology*: e03573.
- 785 Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and
786 Gaussian Markov random fields: the stochastic partial differential equation
787 approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*
788 **73**(4): 423–498. Wiley Online Library.
- 789 Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., and Williams,
790 P.H. 2003. Avoiding pitfalls of using species distribution models in conservation
791 planning. *Conservation biology* **17**(6): 1591–1600.

- 792 Maureaud, A., Frelat, R., Pécuchet, L., Shackell, N., Mérigot, B., Pinsky, M.L., Amador, K.,
793 Anderson, S.C., Arkhipkin, A., Auber, A., Barri, I., Bell, R.J., Belmaker, J., Beukhof, E.,
794 Camara, M.L., Guevara-Carrasco, R., Choi, J., Christensen, H.T., Conner, J., Cubillos,
795 L.A., Diadhiou, H.D., Edelist, D., Emblemsvåg, M., Ernst, B., Fairweather, T.P., Fock,
796 H.O., Friedland, K.D., Garcia, C.B., Gascuel, D., Gislason, H., Goren, M., Guitton, J.,
797 Jouffre, D., Hattab, T., Hidalgo, M., Kathena, J.N., Knuckey, I., Kidé, S.O., Koen-Alonso,
798 M., Koopman, M., Kulik, V., León, J.P., Levitt-Barmats, Y., Lindegren, M., Llope, M.,
799 Massiot-Granier, F., Masski, H., McLean, M., Meissa, B., Mérillet, L., Mihneva, V.,
800 Nunoo, F.K.E., O'Driscoll, R., O'Leary, C.A., Petrova, E., Ramos, J.E., Refes, W.,
801 Román-Marcote, E., Siegstad, H., Sobrino, I., Sólmundsson, J., Sonin, O., Spies, I.,
802 Steingrund, P., Stephenson, F., Stern, N., Tserkova, F., Tserpes, G., Tzanatos, E., van
803 Rijn, I., van Zwieten, P.A.M., Vasilakopoulos, P., Yepsen, D.V., Ziegler, P., and Thorson,
804 J. 2020. Are we ready to track climate-driven shifts in marine species across
805 international boundaries? - A global survey of scientific bottom trawl data. *Global*
806 *Change Biology: gcb.15404*. doi:10.1111/gcb.15404.
- 807 Milisenda, G., Garofalo, G., Fiorentino, F., Colloca, F., Maynou, F., Ligas, A., Musumeci, C.,
808 Bentes, L., Gonçalves, J.M.S., Erzini, K., Russo, T., D'Andrea, L., and Vitale, S. 2021.
809 Identifying Persistent Hot Spot Areas of Undersized Fish and Crustaceans in
810 Southern European Waters: Implication for Fishery Management Under the Discard
811 Ban Regulation. *Frontiers in Marine Science* **8**: 60. doi:10.3389/fmars.2021.610241.
- 812 Moreno, A., Pereira, J., Arvanitidis, C., Robin, J.-P., Koutsoubas, D., Perales-Raya, C., Cunha,
813 M.M., Balguerías, E., and Denis, V. 2002. Biological variation of *Loligo vulgaris*
814 (Cephalopoda: Loliginidae) in the eastern Atlantic and Mediterranean. *Bulletin of*
815 *Marine Science* **71**(1): 515–534.
- 816 Murray, L.G., Hinz, H., Hold, N., and Kaiser, M.J. 2013. The effectiveness of using CPUE data
817 derived from Vessel Monitoring Systems and fisheries logbooks to estimate scallop
818 biomass. *ICES Journal of Marine Science* **70**(7): 1330–1340.
- 819 Nielsen, J.R. 2015. Methods for integrated use of fisheries research survey information in
820 understanding marine fish population ecology and better management advice:
821 improving methods for evaluation of research survey information under
822 consideration of survey fish detection and catch efficiency. Wageningen University.
- 823 Nielsen, J.R., Thunberg, E., Holland, D.S., Schmidt, J.O., Fulton, E.A., Bastardie, F., Punt, A.E.,
824 Allen, I., Bartelings, H., and Bertignac, M. 2018. Integrated ecological–economic
825 fisheries models—Evaluation, review and challenges for implementation. *Fish and*
826 *Fisheries* **19**(1): 1–29.
- 827 Ord, J.K., and Getis, A. 1995. Local Spatial Autocorrelation Statistics: Distributional Issues
828 and an Application. *Geographical Analysis* **27**(4): 286–306. doi:10.1111/j.1538-
829 4632.1995.tb00912.x.
- 830 Pedersen, S.A., Fock, H.O., and Sell, A.F. 2009. Mapping fisheries in the German exclusive
831 economic zone with special reference to offshore Natura 2000 sites. *Marine Policy*
832 **33**(4): 571–590. doi:10.1016/j.marpol.2008.12.007.
- 833 Pennino, M.G., Conesa, D., Lopez-Quilez, A., Munoz, F., Fernández, A., and Bellido, J.M. 2016.
834 Fishery-dependent and-independent data lead to consistent estimations of essential
835 habitats. *ICES Journal of Marine Science* **73**(9): 2302–2310. Oxford University Press.
- 836 Pennino, M.G., Paradinas, I., Illian, J.B., Muñoz, F., Bellido, J.M., López-Quílez, A., and Conesa,
837 D. 2019. Accounting for preferential sampling in species distribution models.
838 *Ecology and evolution* **9**(1): 653–663.

- 839 Petitgas, P. 1997. Sole egg distributions in space and time characterised by a geostatistical
840 model and its estimation variance. *ICES Journal of Marine Science* **54**(2): 213–225.
- 841 Pinto, C., Travers-Trolet, M., Macdonald, J.I., Rivot, E., and Vermard, Y. 2019. Combining
842 multiple data sets to unravel the spatiotemporal dynamics of a data-limited fish
843 stock. *Canadian Journal of Fisheries and Aquatic Sciences* **76**(8): 1338–1349. NRC
844 Research Press.
- 845 Ramzi, A., Arino, O., Koutsikopoulos, C., Boussouar, A., and Lazure, P. 2001. Modelling and
846 numerical simulations of larval migration of the sole (*Solea solea* (L.)) of the Bay of
847 Biscay. Part 2: numerical simulations. *Oceanologica Acta* **24**(2): 113–124.
848 doi:10.1016/S0399-1784(00)01132-4.
- 849 Regimbart, A., Guitton, J., and Le Pape, O. 2018. Zones fonctionnelles pour les ressources
850 halieutiques dans les eaux sous souveraineté française. Deuxième partie : Inventaire.
851 Rapport d'étude. Les publications du Pôle halieutique A. Pôle halieutique
852 AGROCAMPUS OUEST, Rennes.
- 853 Rufener, M.-C., Kristensen, K., Nielsen, J.R., and Bastardie, F. 2021. Bridging the gap between
854 commercial fisheries and survey data to model the spatiotemporal dynamics of
855 marine species. *Ecological Applications*: e02453.
- 856 Salas, S., and Gaertner, D. 2004. The behavioural dynamics of fishers: management
857 implications. *Fish and Fisheries* **5**(2): 153–167. doi:https://doi.org/10.1111/j.1467-
858 2979.2004.00146.x.
- 859 Silvano, R.A., MacCord, P.F., Lima, R.V., and Begossi, A. 2006. When does this fish spawn?
860 Fishermen's local knowledge of migration and reproduction of Brazilian coastal
861 fishes. *Environmental Biology of fishes* **76**(2): 371–386.
- 862 Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E., and Semmens, B.X. 2019. The utility of
863 spatial model-based estimators of unobserved bycatch. *ICES Journal of Marine
864 Science* **76**(1): 255–267. doi:10.1093/icesjms/fsy153.
- 865 Thorson, J.T. 2018. Three problems with the conventional delta-model for biomass
866 sampling data, and a computationally efficient alternative. *Canadian Journal of
867 Fisheries and Aquatic Sciences* **75**(9): 1369–1382. NRC Research Press.
- 868 Thorson, J.T., Adams, C.F., Brooks, E.N., Eisner, L.B., Kimmel, D.G., Legault, C.M., Rogers, L.A.,
869 and Yasumiishi, E.M. 2020a. Seasonal and interannual variation in spatio-temporal
870 models for index standardization and phenology studies. *ICES Journal of Marine
871 Science* **77**(5): 1879–1892.
- 872 Thorson, J.T., Ciannelli, L., and Litzow, M.A. 2020b. Defining indices of ecosystem variability
873 using biological samples of fish communities: A generalization of empirical
874 orthogonal functions. *Progress in Oceanography* **181**: 102244.
875 doi:10.1016/j.pocean.2019.102244.
- 876 Tidd, A.N., Vermard, Y., Marchal, P., Pinnegar, J., Blanchard, J.L., and Milner-Gulland, E.J.
877 2015. Fishing for Space: Fine-Scale Multi-Sector Maritime Activities Influence Fisher
878 Location Choice. *PLOS ONE* **10**(1): e0116335. doi:10.1371/journal.pone.0116335.
- 879 Wikle, C.K., Zammit-Mangion, A., and Cressie, N. 2019. *Spatio-temporal Statistics with R*.
880 CRC Press.
- 881 Yan, Y., Cantoni, E., Field, C., Treble, M., and Flemming, J.M. 2022. Spatiotemporal modeling
882 of bycatch data: methods and a practical guide through a case study in a Canadian
883 Arctic fishery. *Canadian Journal of Fisheries and Aquatic Sciences* **79**(1): 148–158.
- 884 Yochum, N., Starr, R.M., and Wendt, D.E. 2011. Utilizing fishermen knowledge and expertise:
885 keys to success for collaborative fisheries research. *Fisheries* **36**(12): 593–605.

886

887

Draft

888 **TABLES**

889

890 Table 1. Ratio between the negative log-likelihood values (either commercial or
 891 scientific) from the integrated model accounting for preferential sampling and the
 892 integrated model ignoring preferential sampling.

Species	Negative log-likelihood ratio	
	Scientific data	Commercial data
Sole	0.97	0.92
Squids	1.00	1.01
Whiting	0.99	1.00

893

Note. The ratio between negative log-likelihoods ($-\log(lkl)$) is given as: $r = \frac{-\log(lkl_{PS})}{-\log(lkl_{noPS})}$.

894

If $r < 1$, the model accounting for PS better fits the data than the model ignoring PS (no PS).

895

Draft

896 **FIGURES CAPTION**

897 Figure 1. Spatial distribution of each fleet on the whole period (2010-2018). Unit: fishing
898 effort in fishing hour.

899

900 Figure 2. Diagram of the integrated spatio-temporal model.

901

902 Figure 3. Sole case study. Comparison between the observed scientific CPUE (y-axis)
903 and the corresponding model predictions (x-axis) on the month of the survey, based on
904 model integrating data from one commercial fleet only (either OTB_CEP, OTB_DEF,
905 OTT_DEF) or from all commercial fleets (Integrated model). x-axis: model predictions. y-
906 axis: scientific data observations (CPUE in kg/hour). Black line: linear regression
907 'log(science observations) ~ log(model predictions)'. r : Spearman correlation coefficient.
908 Scientific data are integrated to inference for all models. *** stands for the level of
909 significance. Each point is a grid cell in which fall a scientific data point. We compare the
910 scientific observation values to the related prediction values.

911

912 Figure 4. Sole case study. Comparison between predictions (each point is a combination
913 'grid cell x time step'.) from the integrated model (using all fishing fleets) and the model
914 integrating only one commercial fleet for the 12 months of year 2018. Left: OTB_CEP
915 fleet, middle: OTB_DEF fleet, right: OTT_DEF fleet. x-axis: integrated model predictions.
916 y-axis: single-fleet model predictions. The prediction values are log-scaled. Red points:
917 predictions within the sampling area of the related fleets (i.e. the cells sampled by the
918 fleet). Black points: predictions outside the sampling area of the related fleets. Black line:
919 $x = y$ axis. Note that the intercept of the x-y line has been scaled to account for differences
920 in the intercept values between models. Scientific data are integrated to inference for all
921 models.

922

923

924 Figure 5. Estimates of PS parameters for each commercial fleet (left) and effect of PS
925 on model outputs (right). Left: boxplot represent the variability of maximum likelihood
926 estimates of parameters b across all monthly time steps. Right: log-predictions of the
927 integrated model accounting for PS (y-axis) versus log-predictions of the integrated model
928 ignoring PS (x-axis) for the 12 months of year 2018. Blue points: predictions within the
929 sampling area of the commercial fleets (i.e. the cells sampled by commercial fleets). Black
930 point: predictions outside the sampling area of the commercial fleets. Black line: $x = y$
931 axis.

932

933 Figure 6. Sole case study. (Top) Temporal evolution of the b parameters for the three
934 commercial fleets fitted to the integrated model. Blue vertical lines: January. (Bottom)
935 Monthly biomass distribution averaged on the full period. Only quantile values are
936 represented. Model predictions come from the integrated model accounting for PS.

937

938 Figure 7. Left: index of persistence (average over 2010-2018) during the reproduction
939 period of sole (February), whiting (March-May) and squids (January-April). Reproduction
940 period defined from ecological expertise. Right: literature information on reproduction
941 grounds when available. For sole, the map represents egg concentration from an egg and
942 larvae survey conducted in 1982 (Arbault et al., 1986). For whiting, the map represents
943 records of age-2+ whiting (i.e. mature individuals), from two spring trawl surveys that
944 occurred between 1987 and 1992 (House and Forest, 1993). Model predictions come
945 from the integrated model accounting for PS.

946

947 Figure 8. Persistence indices within the reproduction period computed on a 5-years
948 mobile time-span for each 3 species (5-years time span indicated on the top of each map).
949 Model predictions come from the integrated model accounting for PS.

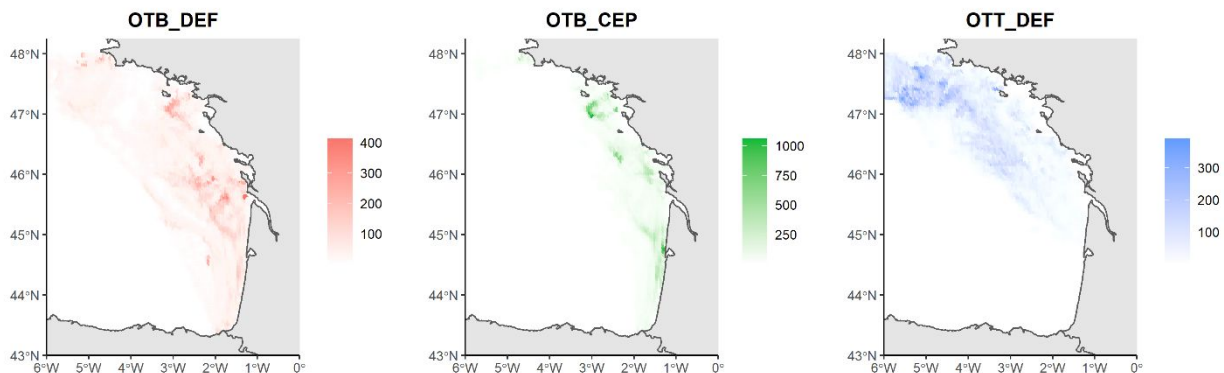
950

951

Draft

952 **FIGURES**

953



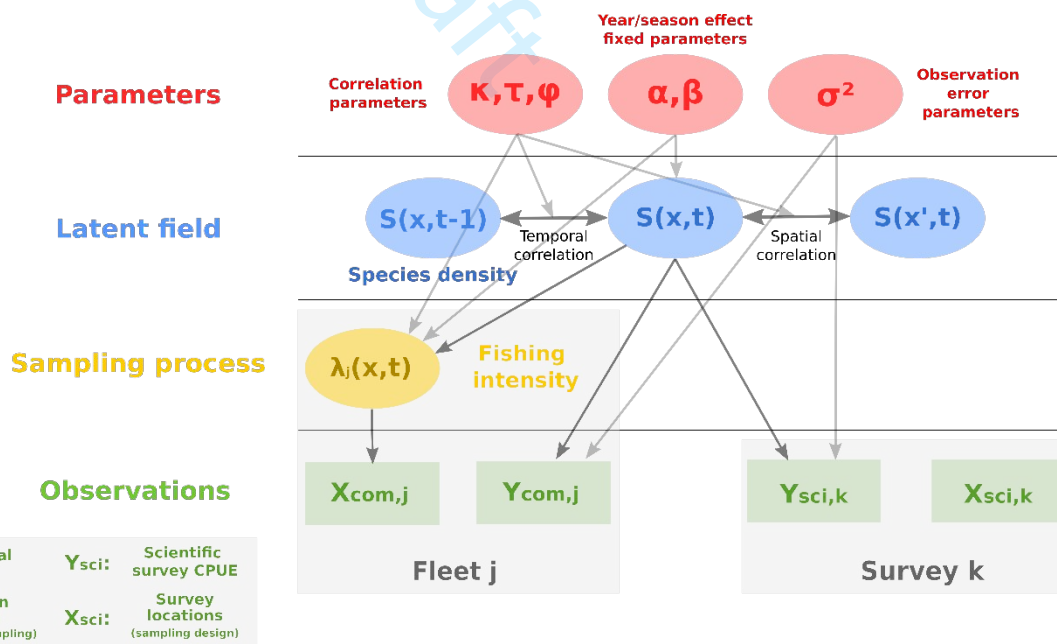
954

955

956

957

Figure 1

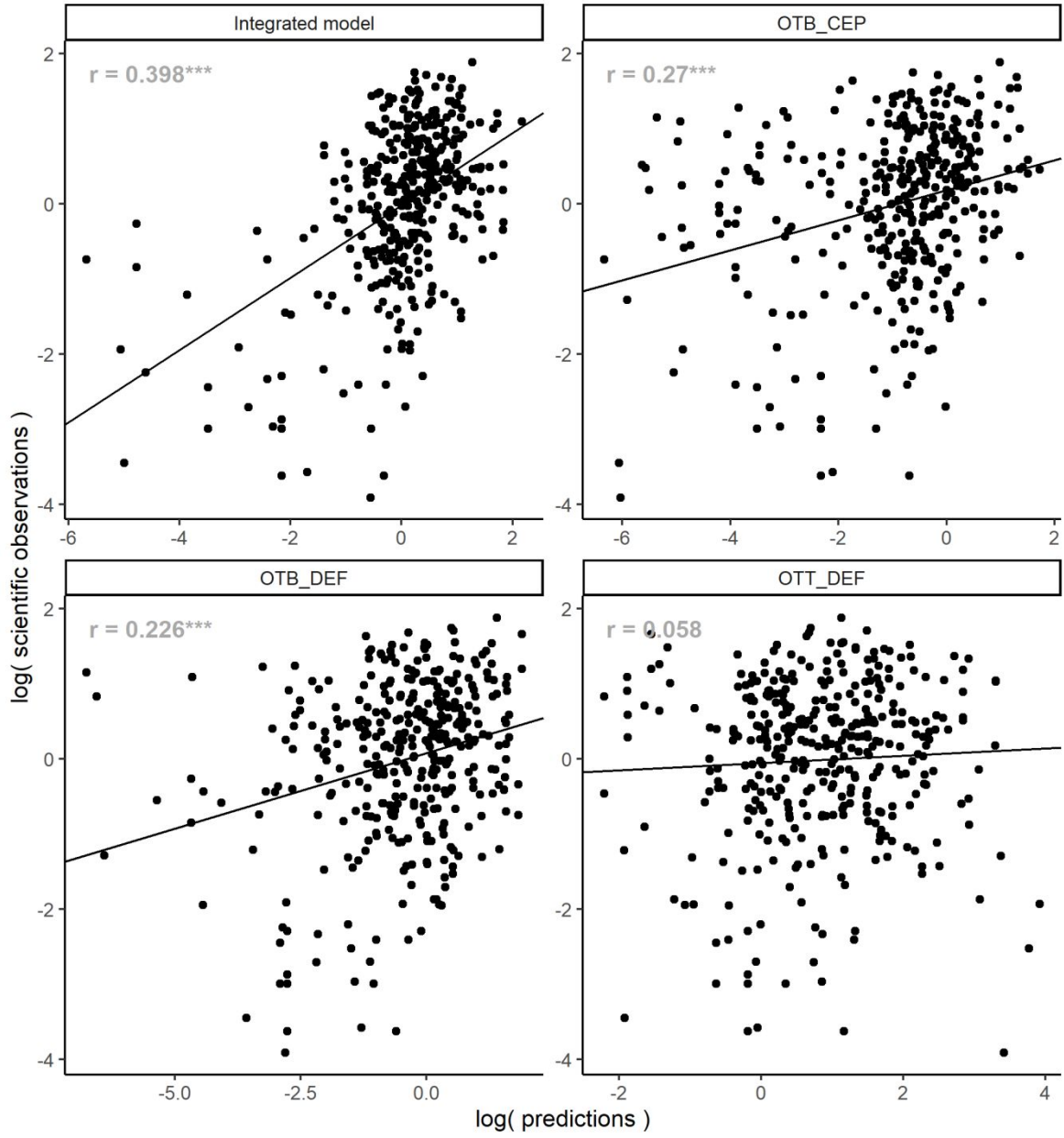


958

959

960

Figure 2



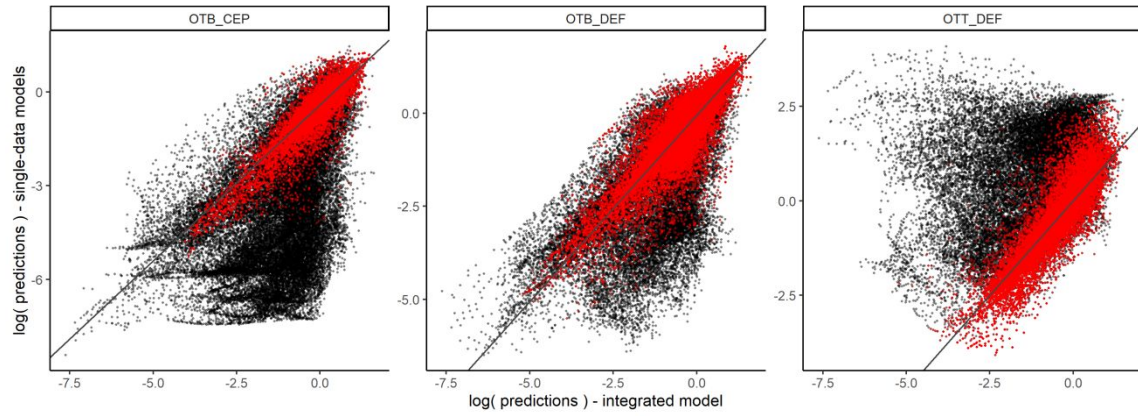
961

962

963

964

Figure 3

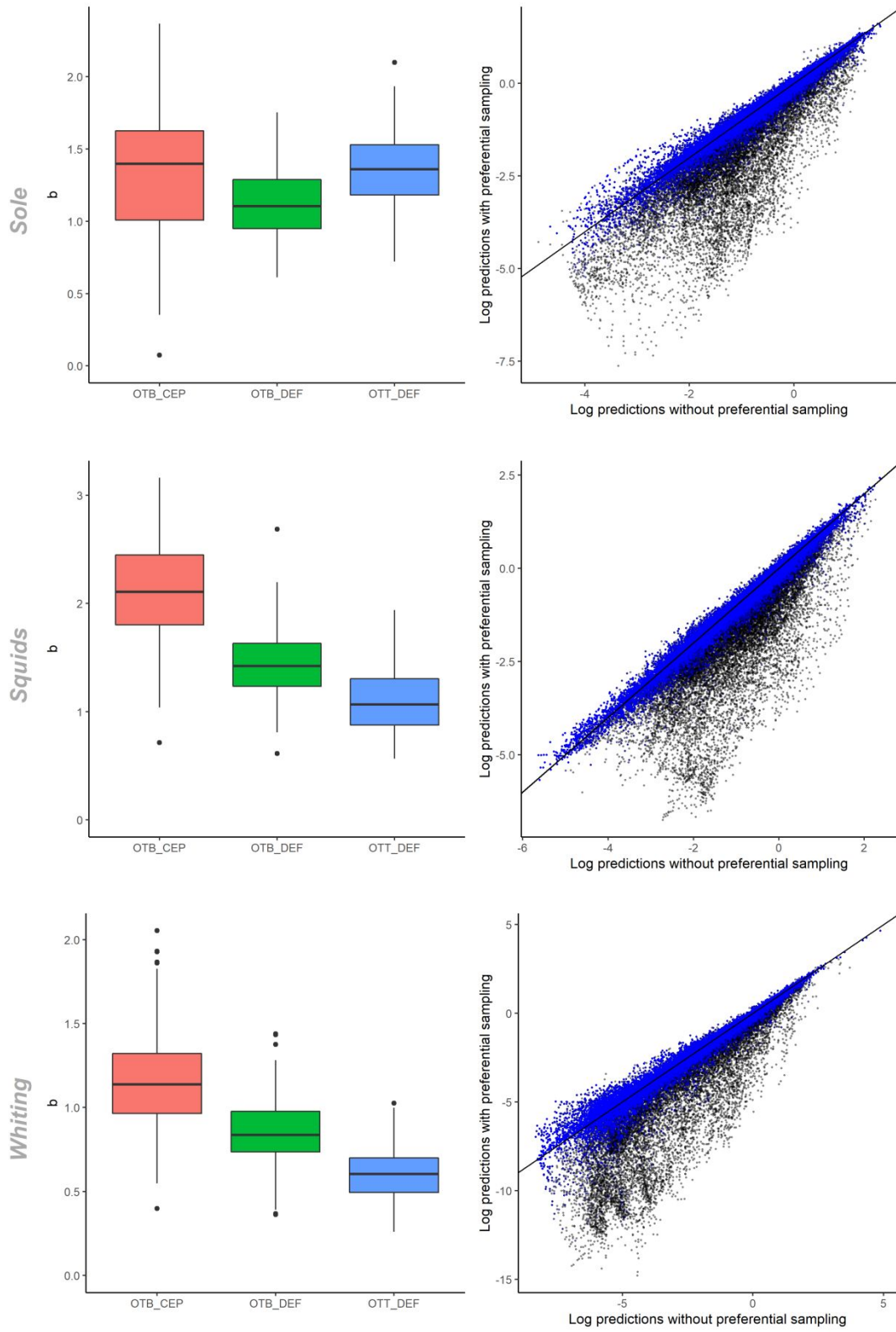


965

966

Figure 4

Draft

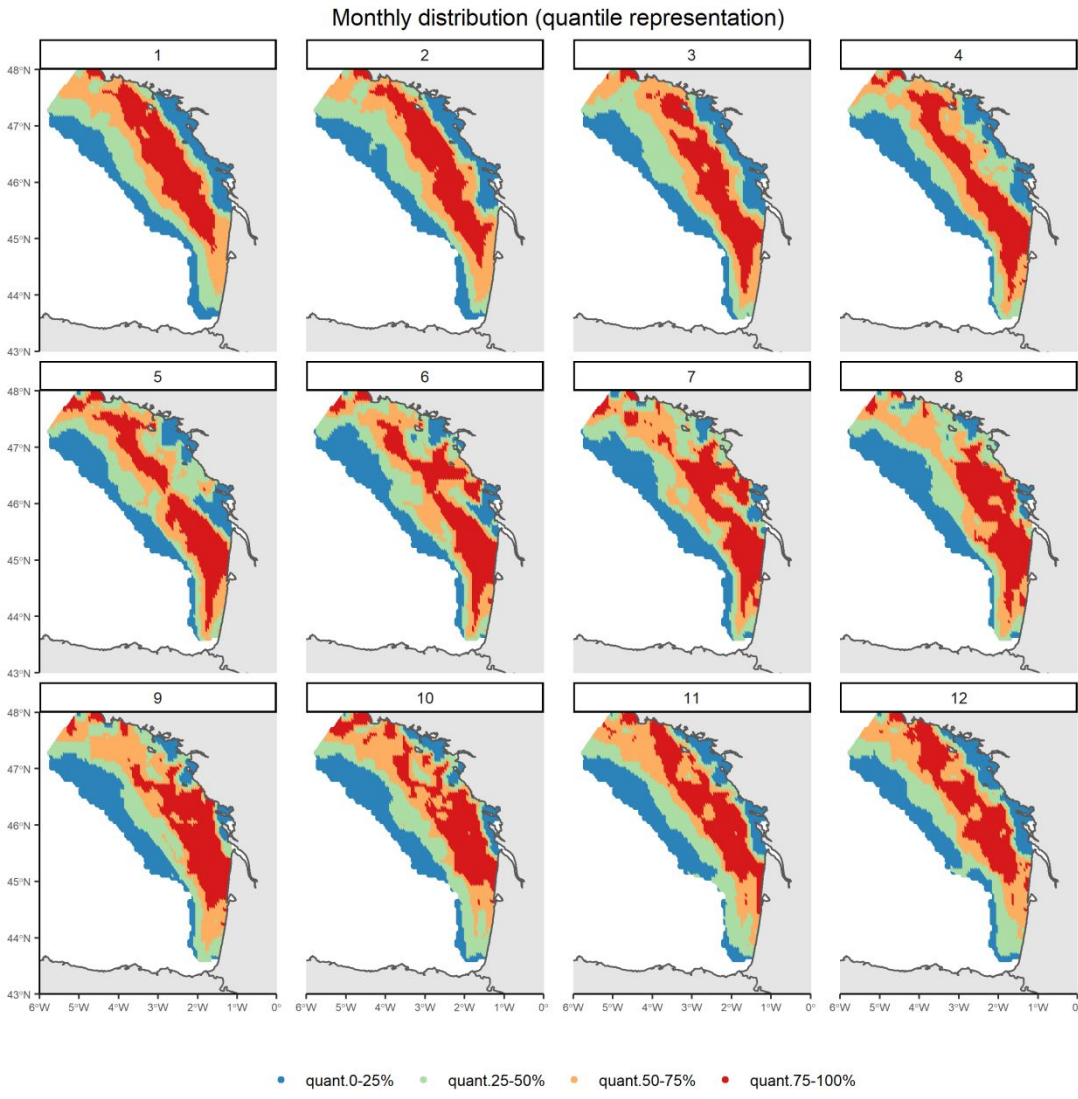
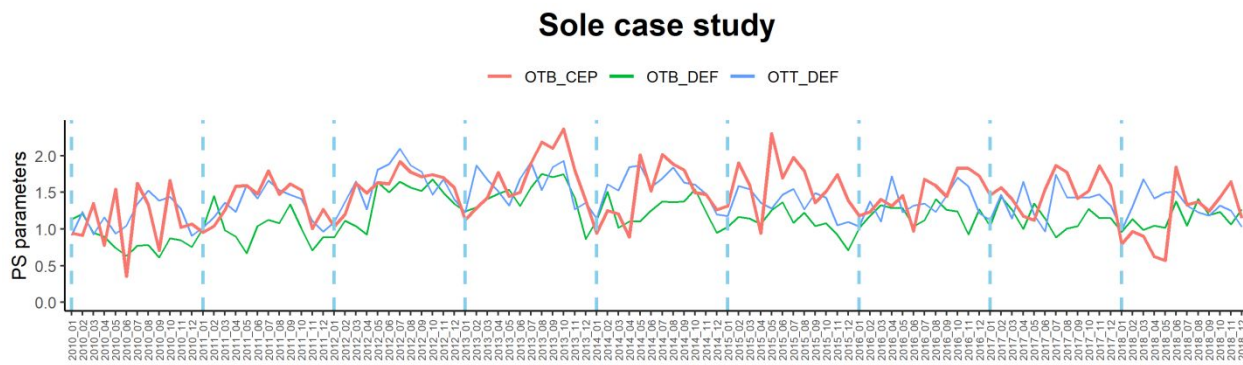


967

968

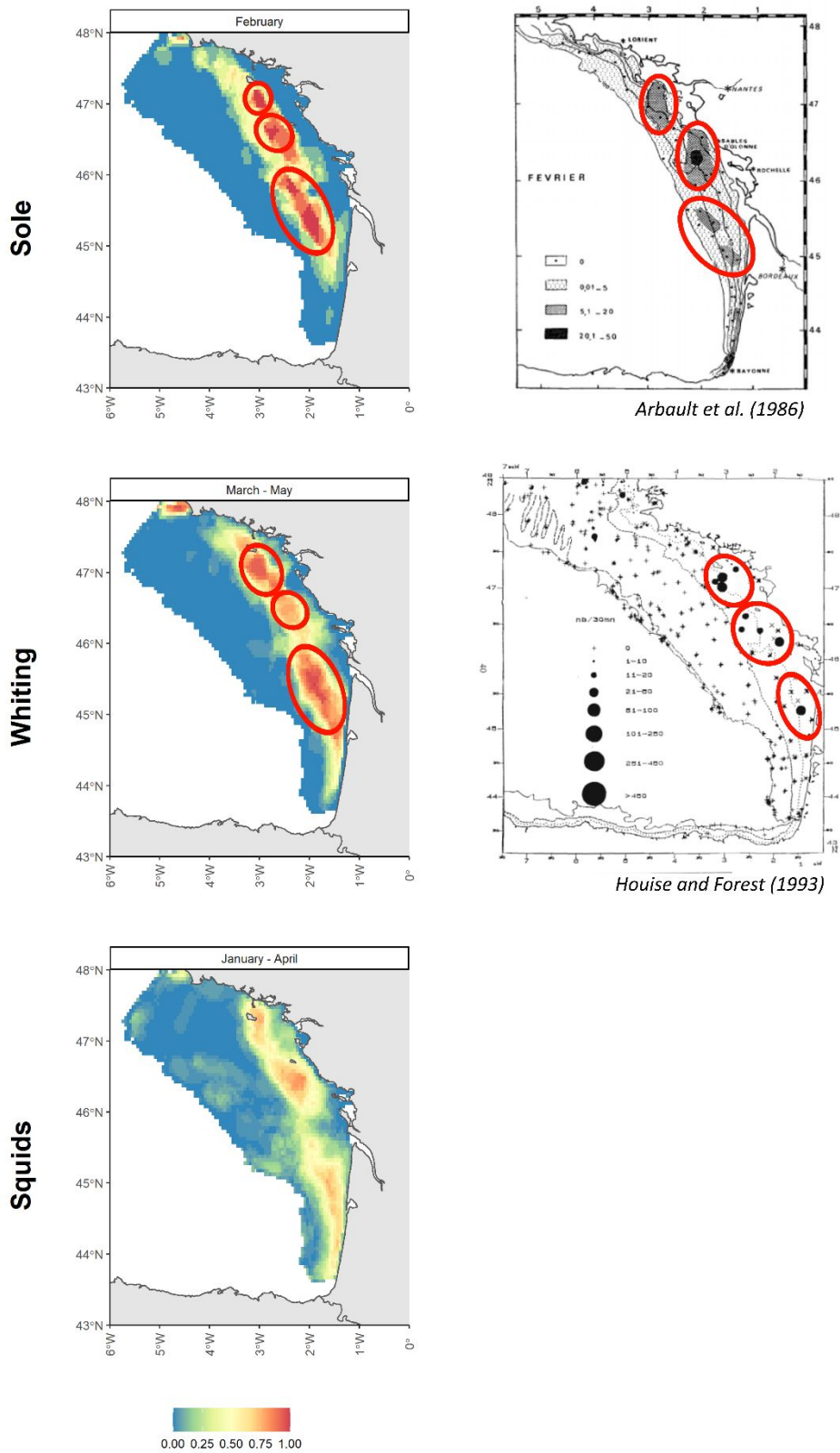
969

Figure 5



970
971
972

Figure 6

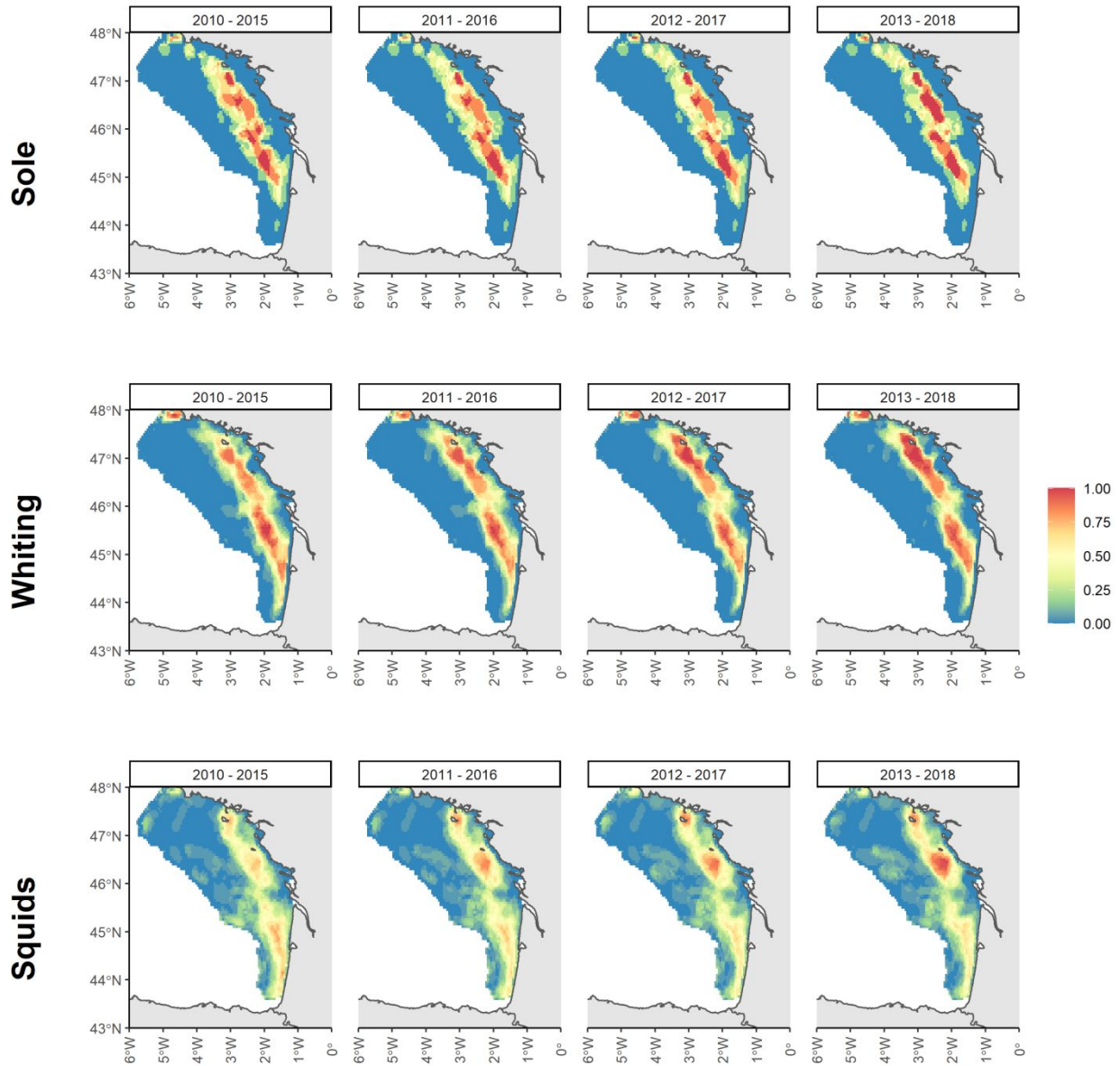


973

974

Figure 7

975



976

977

Figure 8

978 **APPENDICES**

979 The appendices are in the supplementary material file.