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

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

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

Baptiste Alglave[1,2 *], Youen Vermard[1], Etienne Rivot[2], Marie-Pierre Etienne[3], & Mathieu Woillez[4]

[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. Data available through commercial declarations provide valuable information covering the whole year and consequently they could prove useful to identify spawning grounds. 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 fishermen towards areas of higher biomass. Over the spawning period, we applied a method to identify areas of persistent aggregation of biomass and interpret these as spawning areas. The model is applied to infer monthly maps of three species (sole, whiting, squids) in the Bay of Biscay on a 9-years period. Integrating several commercial fleets in inference provide a good coverage of the study area and improves model predictions. The preferential sampling parameters give insights into the temporal dynamics of the targeting behavior of the different fleets. Persistent aggregation areas reveal consistent with the available literature on spawning grounds, highlighting that our approach allows to identify potential areas of reproduction.

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

1 INTRODUCTION

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

The available data to map fish distribution and identify essential habitats mainly rely on either scientific survey data (fishery-independent data) or commercial data available through on-board observer programs (fishery-dependent data) (Pennino et al. 2016). Both data sources benefit of direct on-board recording of catches and are usually considered as high quality data. Furthermore, both data sources were proved to be complementary (Rufener et al. 2021). Scientific data benefit from a standardized sampling plan, a constant catchability and occur each year at the same period. Consequently, they provide standardized data on a large spatial extent for the most species and size classes (Hilborn and Walters 2013; Nielsen 2015). On the other hand, observer data potentially provide data over the full year for all caught species, even though they do not follow a standardized protocol as survey data. Both are characterized by a relatively low sampling intensity in space and time. Because of material limitations, surveys occur only once or twice a year and provide a limited number of sample each time (ICES 2005) and observer programs only cover a limited fraction of the entire fleet (e.g. only 1% of all sea trips are covered by the French observer programs - Cornou et al., 2021). The low sampling density of both data sources may lead to imprecise predictions (ICES, 2005; Alglave et al., 2022) and

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

Commercial catch declarations combined with their fishing locations available from VMS (Vessel Monitoring System) were proven to be an interesting alternative to obtain catch per unit effort (CPUE) data with fine spatial and temporal resolution (Pedersen et al. 2009; Bastardie et al. 2010; Gerritsen and Lordan 2010; Hintzen et al. 2012; Murray et al. 2013; Azevedo and Silva 2020). However, considering commercial fisheries data to infer fish spatial distribution remains highly challenging. Among other challenges, this implies accounting for fishermen sampling behavior. Fishermen typically tend to preferentially sample areas of higher biomass (a process referred to as preferential sampling, PS - Diggle et al. 2010) which can lead to biased spatial predictions if not accounted for in inference (Conn et al. 2017; Pennino et al. 2019; Alglave et al. 2022).

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

In this paper, we extend the modeling framework from Alglave et al. (2022) with a temporal dimension to estimate fish spatio-temporal distribution at a monthly time step. Our new model accounts for the variation over time (monthly time step) in the biomass field as well

as in the intensity of PS of the distinct fishing fleets. To illustrate the value of the method, we selected and applied the model to 3 demersal species in the Bay of Biscay (common sole, whiting and squids) characterized by contrasted configurations regarding the available knowledge of their spawning grounds. We used those applications to reinforce results obtained in Alglave et al. (2022) demonstrating how the integrated framework benefit from the huge amount of spatio-temporal CPUE data to produce accurate maps of spatio-temporal biomass. To illustrate the capacity of the framework to identify potential spawning grounds, we processed model outputs to identify areas of recurrent aggregation occurring during the reproduction season and confronted these to literature information.

2 MATERIAL AND METHODS

In this section, we first present the different species, the datasets and how we process and combine them to produce CPUE data in space and time. Second, we extend the model proposed by Alglave et al. (2022) to introduce a temporal dimension on a discrete monthly time step. Then, we illustrate how integrating several fleets in the analysis improves models predictions, how the PS component modifies model predictions and can be interpreted. Last, we illustrate how we investigate spatio-temporal dynamics from model outputs and identify reproduction grounds based on the aggregation patterns of each 3 species. The models were fitted to data from 2010 to 2018 on a monthly time step (108 time steps).

2.1 Case studies

Sole is a data-rich case. Direct information about reproduction grounds is available through egg and larvae surveys (Arbault et al. 1986). Discard rates is also very low, which makes the landings data a good proxy of the catch (ICES 2019a). Whiting is a data-poor case study where only indirect information of reproduction period exists through spring trawl surveys (Houise and Forest 1993). Discard rates can be high (about 30 %) and thus landings data may provide a biased picture of the real catches (ICES 2019b). Squids represent a mixture of several species: *Loligo Vulgaris* (Lamarck, 1798), *Loligo forbesii* (Steenstrup, 1856) and *Alloteuthis sp* (Lamarck, 1798). They are declared under a common denomination in the catch (Loliginidae here referred as squids). Overall scientific survey suggest that the predominant species in the Bay of Biscay is *Loligo Vulgaris* (ICES, 2020a, p.17). All 3 species are data-poor: no information exist regarding their reproduction grounds but some information of the reproduction period exist for *Loligo Vulgaris* (Moreno et al. 2002).

2.2 Data

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

We pre-processed the VMS and catch declaration (logbook) data to obtain high spatial resolution CPUE data for the mature component of those three stocks, for three different fishing fleet, and for each month of the 2010-2018 time series. In the text, CPUE is used but no discard information is available at the scale of the fishing sequence.

We selected data of three trawlers métiers OTB_DEF (bottom otter trawl targeting demersal species), OTB_CEP (bottom otter trawl targeting cephalopods) and OTT_DEF (multi-rig otter trawl targeting demersal species). Here the term fleets is used to refer to these groups. They refer to distinct component that have overall similar targeting behaviors and similar technical characteristics. These fleets were selected (1) so as to cover the full spatial domain (Figure 1) and (2) because fishing time of trawlers is a good proxy of effort which allows to compute reliable CPUE for biomass (Hovgêrd and Lassen 2008).

Because one of our primarily goal is to identify spawning grounds, we filtered only the mature fraction of the landings. This was done by crossing the landings data with length class and maturity data (see details in SM1). Note that this procedure was not possible for squids, as there are no data on maturity and size classes in this case.

Landing data were then combined with VMS data to finally obtained high spatial resolution CPUE data discretized on a 0.05°x0.05° grid on a monthly time step (see the detailed procedure for this combination in Alglave et al. (2022) and SM2).

2.2.2 Scientific data

We also integrated scientific data in the analysis. For hake and squids we used the survey data from the EVHOE survey. The Orhago survey was used for sole (ICES, 2020 - see SM, Figure S3). The data were extracted from the DATRAS database on the period 2010 - 2018. Only the mature fraction of the survey catches were kept in the analysis to make it comparable with commercial data. Some details on the surveys are given in SM3.

2.3 Spatio-temporal integrated model

We build on the integrated hierarchical model proposed by Alglave et al. (2022) by incorporating a temporal component to model the evolution of the latent field of biomass across the monthly time steps (Figure 2). The PS of commercial data for the different fleets is modelled explicitly through inhomogeneous Poisson point processes for the fishing locations.

2.3.1 Biomass field

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

$$\log\big(S(x,t)\big) = \alpha_S(t) + \delta(x,t) \qquad (1)$$

where $x \in D \subset R^2$ stands for the spatial locations and $t \in [\![1,T]\!]$ the monthly time steps. The term $\alpha_S(t)$ is a time varying intercept modelled as a fixed effect and $\delta(x,t)$ is a GRF spatio-temporal process which represents the spatio-temporal correlation structure of the biomass field. As commercial data may not always cover the full area, the temporal correlation component allows to interpolate between time-steps. Here, the spatio-temporal term has a classical stationary first-order autoregressive form (eq.2) following (Cameletti et al. 2013):

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

155 The autocorrelation coefficient φ is a scalar with $\varphi \in]-1,1[$, $\omega(x,t)$ represents the spatial
156 innovation and is a 0 mean GRF (with no temporal correlation).

157 Note that no covariate is included in the latent field to keep the model as simple as
158 possible. If any, the covariates effects are captured through the spatio-temporal term
159 $\delta(x,t)$. Similarly, the intercept $\alpha_S(t)$ was modelled through a simple fixed effect but more
160 complex specifications including some seasonal, yearly and interaction effect could be
161 adopted such as in Thorson et al. (2020).

162 ***2.3.2 Sampling process for the commercial fishing points***

163 As the scientific survey sampling plan is designed independently from the biomass field,
164 scientific sampling locations do not need to be modelled explicitly (Diggle et al. 2010). By
165 contrasts, the dependence between the fishing locations and the biomass field has to be
166 modelled to capture preferential sampling. We extended the model proposed by Alglave
167 et al. (2022) to account for temporal variations in PS. Fishing locations are modelled as
168 an inhomogeneous point process (X_{comj} in the Figure 2) whose intensity $\lambda_j(x,t)$ (eq.3)
169 controls the expected number of fishing points within a given area:

170 $$log\left(\lambda_j(x,t)\right) = \alpha_{Xj}(t) + b_j(t).log\big(S(x,t)\big) + \eta_j(x,t) \qquad (3)$$

171 where:

172 - the time varying intercept $\alpha_{Xj}(t)$ quantifies the average fishing intensity on the
173 whole area; as the biomass intercept $\alpha_S(t)$, it is modelled as a fixed effect;

174 - the time varying $b_j(t)$ quantifies the strength of PS; it is modelled as a fixed effect
175 too. If $b_j(t) = 0$, then PS is null. If $b_j(t) > 0$, then PS occurs and the greater, the
176 stronger PS;

- the pure spatial GRF $\eta_j(x,t)$ captures the remaining spatial variability in the fishing point pattern not captured by the PS term (for instance, dependence of the fishing locations towards management regulations, distribution of other targeted species, habits/tradition).

2.3.3 Observation process

All observations for both scientific and commercial data of any fleet j are assumed all mutually independent conditionally on the latent field of biomass and the sampling locations. The observation model allows to distinguish several fleets with specific catchabilities. As data (both scientific and commercial) eventually present a high proportion of zero values, we model the observations through a Poisson-link zero-inflated model introduced by Thorson (2018) and already used in Alglave et al. (2022) (see detailed description of the observation model in SM4).

2.3.4 Maximum likelihood estimation

The estimation of the spatio-temporal model is achieved through maximum likelihood estimation. We used the SPDE approach Lindgren et al. (2011) and Template Model Builder (TMB - Kristensen et al., 2016) for a fast estimation of the spatial and spatio-temporal random effects. Details on estimation are provided in SM 5, 6 and 7.

2.4 Evaluating the Interest of integrating multiple fleets

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

between scientific data and model predictions. We also compared the coefficient of variation of the prediction between each model (for November 2018).

2.5 Evaluating the value of modelling PS

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

2.6 Interpreting the intensity of preferential sampling

The estimates of PS parameters may bring valuable information on the dynamics of the fishery as they inform on the strength of the relationship between commercial sampling distribution and species distribution. We investigate the variability of the PS parameters (b) by representing the variability of the different b parameter estimates for the three case studies and the different fleets. Then, focusing on the sole case study, we highlight the insights brought by the model on the temporal evolution of PS and its seasonal variations.

2.7 Investigating spatio-temporal dynamics and identifying reproduction grounds

The spatio-temporal model provides some insight on the temporal dynamics of species distribution both at inter- and intra-annual levels. Based on the maps of abundance inferred at each time steps, we applied a method to identify recurrent aggregation areas.

2.7.1 Aggregation index

We used the Getis and Ord index $G_d\ (x,t)$ (Getis and Ord 1992; Ord and Getis 1995) to determine persistent aggregation areas (see for instance Milisenda et al., (2021)). The generalized version of the Getis and Ord index is given in Bivand and Wong (2018) and Ord and Getis (1995). Basically, $G_d\ (x,t)$ is a normalized version of the ratio between the sum of the log-biomass (denoted $s(x,t)$) within a fixed neighborhood d and the sum of $s(x,t)$ on the entire area (for a fixed time step) (Getis and Ord 1992). We computed these indices on $s(x,t) = \log\big(S(x,t)\big)$ so that the $s(x,t)$ are Gaussian, which makes G_d Gaussian too. In the application, we used a neighborhood distance *d*=7.5 km which defines a small neighborhood of 8 cells (the direct neighbors of each cell grid) and allows to identify very localized aggregation areas. Positive values for the aggregation index G_d indicates that $s(x,t)$ fall within a local patch of high values while negative G_d indicates that $s(x,t)$ fall within a local patch of low values. Near 0 values G_d , indicates that $s(x,t)$ does not fall in some local aggregation patch. As G_d follows a standardized Gaussian distribution, the comparison between the value of the index and the quantiles of a standard Gaussian distribution can be used to evaluate whether or not the latent field of biomass fall within a statistically significant high or low aggregation patch. We used the quantile 99% (2.58) as a threshold to ensure a high level of significance for patch detection (only local patch of positive values are considered) and applied the Bonferroni correction to account for the multiple statistical tests that are conducted.

Then, we define the persistence indices $IP(x,m)$ as the proportion for which a point x falls significantly within an aggregation area for a specific month/season m (can be either a month or several months) among several years. IP allows to define the persistent

aggregation areas during reproduction throughout the time series. They are used in the following to identify reproduction grounds.

2.7.2 Confronting the results with the available literature

We compare the information available from the literature (for sole and whiting) with the persistence indices on their reproduction periods.

Arbault et al. (1986) investigated the reproduction of sole along the Bay of Biscay based on several egg surveys occurring in 1982. Five surveys were conducted between January and May. Egg density was sampled in different locations from Hendaye to Pointe du Raz (43°30N-48°N) and allowed to map the distribution of egg production on the full study domain. The pic of reproduction occurred in February; thus we compare the maps obtained from the February survey with the persistence index we obtained for February.

For whiting, only two EVHOE trawl surveys occurred during spring (considered as the reproduction period of whiting) between 1987 and 1992 in the Bay of Biscay (Houise and Forest 1993). For each haul, the individuals were counted and aged. Individual up to two years were all considered mature. We compare the distribution of age-2 individuals obtained with these surveys (there were very few age 3-individuals) and the index of persistence from our model during spring (March to May).

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

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

3 RESULTS

3.1 Assessing the contribution of each data sources to inference

Results highlight how combining several commercial fleets in the framework brings a better picture of the spatial distribution on the whole domain. For instance, when comparing model prediction the month of the survey and survey outputs, integrating several fleets into the analysis improves correlation with scientific data (Figure 3). It also allows to reduce the standard deviation of the predictions on the full domain (Figure S8).

When looking at the predictions within the spatial range of the fleets, single-fleet models logically provide similar spatial predictions compared with the integrated model (Figure 4; red dots). However, predictions realized outside the spatial range of the fleet largely depart from the ones realized through the integrated models (black dots, Figure 4), emphasizing that the other fleets bring additional information to inference in these areas. This is particularly evidenced with the OTB_CEP and OTT_DEF fleets that partially cover the study area compared with OTB_DEF that better cover the whole study area (Figure 1).

3.2 Interpreting estimates PS intensity

Estimates of the PS intensity (b parameters) for the different species, the different fleets and the different time steps provide information on the targeting behavior which are consistent with expertise. Estimates of b are positive for each species and each fleet (Figure 5, left column). For squids, PS is the strongest for OTB_CEP followed by OTB_DEF and OTT_DEF. This is consistent with the expert knowledge of the targeting behavior of these fleets: OTB_CEP target cephalopods and catch on average 15% of squids while OTB_DEF and OTT_DEF catch respectively 5% and 1% of squid). A similar pattern can be identified for whiting ($b_{OTB\_CEP} > b_{OTB\_DEF} > b_{OTT\_DEF}$); this is consistent with species spatial distribution as whiting (like squids) are found in coastal areas where

the OTB_CEP fleet is operating. For sole, the strength of PS is on average higher for OTB_CEP and OTT_DEF than for OTB_DEF but with less contrast between the three commercial fleets.

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

3.3 Evaluating the influence of PS on spatial distribution

Because estimates of b are positive, spatial density of fishing points is positively correlated with biomass density. Hence, considering PS revises downwards the biomass estimates in areas not sampled by the commercial fleets compared to estimates obtained while ignoring PS (Figure 5, right column, black points), but does not strongly affect predictions in locations within the range of the fleets (blue points). Considering PS only slightly improves the fit of the model to the data. For the Sole case study, some improvement of the likelihood occurred in both the commercial and the scientific likelihood values (Table 1). For whiting and squids, there are no strong modifications in both scientific and commercial likelihoods.

3.4 Investigating spatio-temporal dynamics of fish biomass

Results provide biomass density maps on a monthly time step that allow for evaluating seasonal distribution patterns, and from which aggregation index were calculated. The temporal correlation parameter (φ) is estimated around 0.8 for all the species emphasizing strong temporal correlations in the biomass field values. The range parameters are

318 estimated to 55 km for sole and squids while being estimated to 67 km for whiting
319 emphasizing wider spatial autocorrelation for this species.

320 Concerning the Sole case study, model predictions highlight the relatively offshore
321 distribution from November to April and its coastal distribution from June to October
322 suggesting some migrations happen between these 2 periods (Figure 6, bottom). In
323 particular, the migration in June/July conducts to a contraction of the sole distribution
324 around the Vendée coast, the Gironde Estuary and the Landes coast (45.5°N-46°N) while
325 the migration in November leads to an expansion of the species distribution towards the
326 offshore areas all along the Bay of Biscay. Interestingly such seasonality coincides with
327 the seasonality of PS intensity for the OTB_CEP (Figure 6, Top). Higher PS parameters
328 corresponds to coastal distribution of sole while lower PS parameters corresponds to
329 offshore distribution of sole.

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

331 **3.5 Aggregation index and reproduction grounds**

332 Regarding spawning grounds, both sole and squids emphasize strong aggregation
333 patterns that match the available knowledge of their reproduction grounds. For sole, the
334 aggregation areas globally match with the observed area of maximum egg concentration
335 (Figure 7), although reproduction grounds are slightly eastern in the case of egg maps.
336 This slight discrepancy could be interpreted as an effect of the larval drift as the maps
337 provided by Arbault et al. (1986) are concentration of eggs and not reproduction grounds
338 per se. Overall, these aggregation areas are stable over time (Figure 8).

339 For whiting, similar patterns can be identified during the reproduction period (Figure 7);
340 they match with previous studies investigating the spatial distribution of mature whitings

(Houise and Forest 1993). In particular, the Northern (3°W-47°N) and the Southern (2°W-45.5°N) aggregation patches are almost systematically significantly considered as aggregation areas (aggregation index equals 1) while the other middle one (2.5°W-46.5°N) is classified as an aggregation area that appears less frequently. An additional persistent aggregation area can be identified in the North of the Bay of Biscay (4.5°W-48°N) suggesting that reproduction may also occur in this area which was not identified in the report of Houise and Forest (1993). Interestingly the Northern aggregation area (3°W-47°N) is more pronounced at the end of the period (Figure 8).

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

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

4 DISCUSSION

Main findings

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

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

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

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

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

Combining our results with other data sources to refine inferences on spawning grounds

Although the mature fraction of the biomass was filtered in the data, our maps do not directly inform whether individuals are actually reproducing or not. Persistent aggregation areas should be considered as potential spawning areas rather than actual spawning areas. To get information on actual reproduction zones and support our findings, other kind of data could complement our analysis. Typically, our maps could be of great help to design surveys recording eggs, larvae and spawning individuals that would provide direct information of species reproduction (Fox et al. 2008). Because developing such additional surveys would be highly expensive, our maps could provide valuable a priori information to optimize the survey design and potentially find a compromise between the cost, the spatial extent, the temporal coverage of the survey and the accuracy of the expected estimates/predictions. Similar ideas were already applied to the sole case study to investigate more precisely the space-time variation of sole reproduction. Arbault et al. (1986) work provided a priori information of reproduction grounds that allowed designing more localized surveys to study inter- and intra-annual variability of one specific sole spawning area (Petitgas 1997). Several statistical methods have been developed since and are suitable to optimize such adaptive sampling design; see for instance the recent work of Leach et al. (2021).

Our results could also be combined with fishermen expert knowledge (Yochum et al. 2011) to complement our knowledge of fish reproduction (Delage and Le Pape 2016). For instance, Bezerra et al. (2021) and Silvano et al. (2006) proved the usefulness of fishers knowledge to determine the temporality of fish spawning and to identify some spawning grounds by crossing the information of aggregation areas provided by several fishermen.

These were proved complementary with scientific data as they can be available at low cost and provide local knowledge of fish ecology.

Limits and perspectives for the approach

Our framework has several limitations that are all material for future research avenues.

First, our model remains relatively simple in regards to all the temporal processes that actually occur within a fishery. It is both a strength and a weakness: such way the model remains relatively generic, but one might want to extend it further to account for other temporal and spatio-temporal processes affecting fisheries dynamics. For instance, we opted for a non-seasonal representation of the model, however one could make it seasonal by decomposing the intercepts $\alpha_S(t)$ and $\alpha_{X_j}(t)$ as well as the random effects $\delta(x,t)$ and $\eta(x,t)$ into seasonal and yearly terms in addition to some 'season x year' interaction terms as is performed in Thorson et al. (2020). In their work, such specification mainly allowed to provide information over the time-steps where data was lacking. In the configuration of our case studies, data is available for all time steps and have a relatively good coverage of the study domain. Consequently, such model specification should not modify the overall inference of the biomass field we obtain even though it provide a nice conceptual view of seasonality. Alternatively, our framework could integrate orthogonal spatio-temporal terms in the latent field to capture the main mode of variability of the biomass field (Thorson et al. 2020b). Such orthogonal terms would allow to capture the main spatial patterns that structure the latent field as well as their variation in time. These could prove very useful to identify the structuring processes that affect species distribution and could give a valuable insight in the space-time dynamics of the species. Another exciting research avenue would consist in integrating population dynamics in the latent

455 field of biomass (Cao et al. 2020). This would require to refine further the demographic
456 resolution of the VMS-logbooks data (see for instance Azevedo and Silva 2020), but once
457 done, it would give access to huge data for inferring the space-time dynamics of fish
458 populations. Finally, our model considers fishermen preferentially sample areas where the
459 biomass is higher (preferential sampling), but does not consider any other drivers and
460 specifically the temporal and spatio-temporal relations that can affect fishers behavior.
461 These can be highly complex and may depend on the distribution of the resource,
462 tradition/habits, management regulations (Abbott et al., 2015; Girardin et al., 2017; Salas
463 and Gaertner, 2004; Hintzen, 2021). These drivers are rarely studied in both space and
464 time (although see Tidd et al., 2015). Our framework could allow to jointly model the
465 dynamics of the species, the distribution of the effort, the link that relates species
466 distribution and effort in space/time and all the other spatial and/or temporal drivers that
467 affect the distribution of fishing effort. For instance, we could relate the fishing intensity to
468 the biomass field from the previous time steps, or alternatively consider that the locational
469 choice depend on the catches of the previous time steps. Adding such covariates and
470 spatio-temporal dependencies in the sampling equation (eq.3) will probably not modify the
471 overall pattern of biomass distribution, but it would allow to quantify the drivers of
472 fishermen behavior and give valuable insight in the functioning of the fishery.

473 Second, including discards would potentially improve our approach. Indeed, logbooks
474 data are landings declarations data which means they inform on the landings and not on
475 the true catch. Thus, by assuming the observations we derive from logbooks data are
476 representative of the biomass, we make the hypothesis that the discard rate is constant
477 in space and time and does not affect model predictions. This should not be a problem for
478 sole and squids as the discards are low and TAC have not been really binding during the

studied period. However, the issue might be more stringent for whiting and/or other species with a high and non stationary level of discards. Integrating discards data in the analysis could help solving this issue. Stock et al. (2019) and Yan et al. (2022) used observer data to model bycatch in both space and time and Breivik et al. (2017) used bycatch data from onboard surveys to predict the temporal evolution of bycatch realized in the full commercial data. Similarly, we could integrate into the same analysis the logbooks and the observer data by assuming that the catch of observer data is a summation of both landings (which is also observed in the logbooks data) and discards (which is unobserved in the logbooks data). This way, the discards information available from observer data would be shared with the logbooks data and would allow correcting for the missing portion of catch declarations data while possibly accounting for possible space or time variation in the discard rate.

Last, our analysis rely on the hypothesis that the spawning season is known a priori. Extending the approach to infer the spawning season based on the temporal dynamic of the aggregation patterns could improve our knowledge of species spatio-temporal distribution. In particular, identifying the main species phenomenological phases and their consistency (or shift) in time is crucial in the context of global change (Thorson et al. 2020a).

In our study, we computed the aggregation index on a period we assumed to be the reproduction period based on literature (Arbault et al. 1986; Houise and Forest 1993; Moreno et al. 2002). Hence, our results are sensitive to this a priori hypothesis, or our approach can even be inapplicable for species when no information is available in the literature. Several methods exist and could be adapted to extract the spatial patterns that shape model outputs, their related temporal variation and identify from these the main

phenological phases that characterize species distribution (e.g. reproduction, feeding - see for instance Empirical Orthogonal Functions or Principal Oscillation Patterns - Cressie and Wikle 2015; Wikle et al. 2019).

Future use for Marine Spatial Planning

Finally, our approach could reveal useful in the context of Marine Spatial Planning (MSP). Janßen et al. (2018) highlighted that one of the main requirement for implementing MSP while accounting for fish ecology is the availability of fine scale information on species distribution and of their essential habitats. Here we propose a method which can provide such information for the fraction of the population available through catch declarations (i.e. mainly the adult fraction and in some cases part of the juvenile fraction). This knowledge is typically needed to design Marine Protected Areas (MPA – see for instance Lambert et al. (2017) or Loiselle et al. (2003)), Fishery Conservation Zones (Delage et Le Pape, 2016 ; Regimbart et al., 2018), or alternatively identify areas that should be kept for fishing in a context were many other human activities are competing in space and time with fishing (Campbell et al. 2014; Bastardie et al. 2015). This would require to open the spectrum of the analysis to an economical dimension and possibly integrate our results into ecological–economic models in order to evaluate alternative management regulations and assess their tradeoffs in regards to all the sets of ecosystem services provided through activities such as fishing, aquaculture, energy, shipping, recreation and conservation (Nielsen et al. 2018).

ACKNOWLEDGMENT

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

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

DATA AVAILABILITY STATEMENT

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

REFERENCES

Abbott, J., Haynie, A., and Reimer, M. 2015. Hidden Flexibility: Institutions, Incentives, and the Margins of Selectivity in Fishing. Land Economics **91**: 169–195. doi:10.3368/le.91.1.169.

Alglave, B., Rivot, E., Etienne, M.-P., Woillez, M., Thorson, J.T., and Vermard, Y. 2022. Combining scientific survey and commercial catch data to map fish distribution. ICES Journal of Marine Science: fsac032. doi:10.1093/icesjms/fsac032.

Arbault, P.S., Camus, P., and Bec, C. le. 1986. Estimation du stock de sole (Solea vulgaris, Quensel 1806) dans le Golfe de Gascogne à partir de la production d'œufs. Journal of Applied Ichthyology **2**(4): 145–156. doi:10.1111/j.1439-0426.1986.tb00656.x.

Azevedo, M., and Silva, C. 2020. A framework to investigate fishery dynamics and species size and age spatio-temporal distribution patterns based on daily resolution data: a

case study using Northeast Atlantic horse mackerel. ICES Journal of Marine Science **77**(7–8): 2933–2944. doi:10.1093/icesjms/fsaa170.

Bastardie, F., Nielsen, J.R., Eigaard, O.R., Fock, H.O., Jonsson, P., and Bartolino, V. 2015. Competition for marine space: modelling the Baltic Sea fisheries and effort displacement under spatial restrictions. ICES Journal of Marine Science **72**(3): 824–840. doi:10.1093/icesjms/fsu215.

Bastardie, F., Nielsen, J.R., Ulrich, C., Egekvist, J., and Degel, H. 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fisheries Research **106**(1): 41–53.

Bezerra, I.M., Hostim-Silva, M., Teixeira, J.L.S., Hackradt, C.W., Félix-Hackradt, F.C., and Schiavetti, A. 2021. Spatial and temporal patterns of spawning aggregations of fish from the Epinephelidae and Lutjanidae families: An analysis by the local ecological knowledge of fishermen in the Tropical Southwestern Atlantic. Fisheries Research **239**: 105937.

Biggs, C.R., Heyman, W.D., Farmer, N.A., Kobara, S., Bolser, D.G., Robinson, J., Lowerre-Barbieri, S.K., and Erisman, B.E. 2021. The importance of spawning behavior in understanding the vulnerability of exploited marine fishes in the US Gulf of Mexico. PeerJ **9**: e11814.

Bivand, R.S., and Wong, D.W.S. 2018. Comparing implementations of global and local indicators of spatial association. TEST **27**(3): 716–748. doi:10.1007/s11749-018-0599-x.

Bourdaud, P., Travers-Trolet, M., Vermard, Y., and Marchal, P. 2019. Improving the interpretation of fishing effort and pressures in mixed fisheries using spatial overlap metrics. Can. J. Fish. Aquat. Sci. **76**(4): 586–596. doi:10.1139/cjfas-2017-0529.

Breivik, O.N., Storvik, G., and Nedreaas, K. 2017. Latent Gaussian models to predict historical bycatch in commercial fishery. Fisheries Research **185**: 62–72.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Adv Stat Anal **97**(2): 109–131. doi:10.1007/s10182-012-0196-3.

Campbell, M.S., Stehfest, K.M., Votier, S.C., and Hall-Spencer, J.M. 2014. Mapping fisheries for marine spatial planning: Gear-specific vessel monitoring system (VMS), marine conservation and offshore renewable energy. Marine Policy **45**: 293–300. doi:10.1016/j.marpol.2013.09.015.

Cao, J., Thorson, J.T., Punt, A.E., and Szuwalski, C. 2020. A novel spatiotemporal stock assessment framework to better address fine-scale species distributions: development and simulation testing. Fish and Fisheries **21**(2): 350–367.

Conn, P.B., Thorson, J.T., and Johnson, D.S. 2017. Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. Methods in Ecology and Evolution **8**(11): 1535–1546.

Cornou, A.-S., Quinio-Scavinner, M., Sagan, J., Cloâtre, T., Dubroca, L., and Billet, N. 2021. Captures et rejets des métiers de pêche francais - Résultats des observations à bord des navires de pêche professionnelle en 2019. Ifremer.

Cressie, N., and Wikle, C.K. 2015. Statistics for spatio-temporal data. John Wiley & Sons.

Delage, N., and Le Pape, O. 2016. Inventaire des zones fonctionnelles pour les ressources halieutiques dans les eaux sous souveraineté française. Première partie: Définitions, critères d'importance et méthode pour déterminer des zones d'importance à

protéger en priorité. Rapport de recherche, Pôle halieutique AGROCAMPUS OUEST, Rennes.
Diggle, P.J., Menezes, R., and Su, T. 2010. Geostatistical inference under preferential sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics) **59**(2): 191–232.
Fox, C.J., Taylor, M., Dickey-Collas, M., Fossum, P., Kraus, G., Rohlf, N., Munk, P., van Damme, C.J., Bolle, L.J., and Maxwell, D.L. 2008. Mapping the spawning grounds of North Sea cod (Gadus morhua) by direct and indirect means. Proceedings of the Royal Society B: Biological Sciences **275**(1642): 1543–1548.
Gerritsen, H., and Lordan, C. 2010. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. ICES Journal of Marine Science **68**(1): 245–252.
Getis, A., and Ord, J. 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis.
Girardin, R., Hamon, K.G., Pinnegar, J., Poos, J.J., Thébaud, O., Tidd, A., Vermard, Y., and Marchal, P. 2017. Thirty years of fleet dynamics modelling using discrete-choice models: What have we learned? Fish and Fisheries **18**(4): 638–655. doi.org/10.1111/faf.12194.
Hilborn, R., and Walters, C.J. 2013. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Springer Science & Business Media.
Hintzen, N.T., Aarts, G., Poos, J.J., Van der Reijden, K.J., and Rijnsdorp, A.D. 2021. Quantifying habitat preference of bottom trawling gear. ICES Journal of Marine Science **78**(1): 172–184.
Hintzen, N.T., Bastardie, F., Beare, D., Piet, G.J., Ulrich, C., Deporte, N., Egekvist, J., and Degel, H. 2012. VMStools: Open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data. Fisheries Research **115**: 31–43. Elsevier.
Houise, C., and Forest, A. 1993. Etude de la population du Merlan (Merlangius merlangius L.) du Golfe de Gascogne. Ifremer.
Hovgêrd, H., and Lassen, H. 2008. Manual on estimation of selectivity for gillnet and longline gears in abundance surveys. Food & Agriculture Org.
ICES. 2005. Report of the Workshop on Survey Design and Data Analysis (WKSAD). Sète, France.
ICES. 2019a. Sole (Solea solea) in divisions 8.a–b (northern and central Bay of Biscay). Advice. Available from https://doi.org/10.17895/ices.advice.4775.
ICES. 2019b. Whiting (Merlangius merlangus) in Subarea 8 and Division 9.a (Bay of Biscay and Atlantic Iberian waters). doi:10.17895/ICES.ADVICE.4777.
ICES. 2020a. Working Group on Cephalopod Fisheries and Life History (WGCEPH). ICES. doi:10.17895/ICES.PUB.6032.
ICES. 2020b. International Bottom Trawl Survey Working Group (IBTSWG). ICES Scientific Reports, ICES. Available from http://www.ices.dk/sites/pub/Publication Reports/Forms/DispForm.aspx?ID=37066 [accessed 28 May 2021].
Janßen, H., Bastardie, F., Eero, M., Hamon, K.G., Hinrichsen, H.-H., Marchal, P., Nielsen, J.R., Le Pape, O., Schulze, T., and Simons, S. 2018. Integration of fisheries into marine spatial planning: Quo vadis? Estuarine, Coastal and Shelf Science **201**: 105–113.
Kai, M., Thorson, J.T., Piner, K.R., and Maunder, M.N. 2017. Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (Isurus

oxyrinchus) in the Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences **74**(11): 1765–1780.
Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software **70**(1): 1–21. doi:10.18637/jss.v070.i05.
Lambert, C., Virgili, A., Pettex, E., Delavenne, J., Toison, V., Blanck, A., and Ridoux, V. 2017. Habitat modelling predictions highlight seasonal relevance of Marine Protected Areas for marine megafauna. Deep Sea Research Part II: Topical Studies in Oceanography **141**: 262–274.
Leach, C.B., Williams, P.J., Eisaguirre, J.M., Womble, J.N., Bower, M.R., and Hooten, M.B. 2021. Recursive Bayesian computation facilitates adaptive optimal design in ecological studies. Ecology: e03573.
Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73**(4): 423–498. Wiley Online Library.
Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., and Williams, P.H. 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation biology **17**(6): 1591–1600.
Milisenda, G., Garofalo, G., Fiorentino, F., Colloca, F., Maynou, F., Ligas, A., Musumeci, C., Bentes, L., Gonçalves, J.M.S., Erzini, K., Russo, T., D'Andrea, L., and Vitale, S. 2021. Identifying Persistent Hot Spot Areas of Undersized Fish and Crustaceans in Southern European Waters: Implication for Fishery Management Under the Discard Ban Regulation. Frontiers in Marine Science **8**: 60. doi:10.3389/fmars.2021.610241.
Moreno, A., Pereira, J., Arvanitidis, C., Robin, J.-P., Koutsoubas, D., Perales-Raya, C., Cunha, M.M., Balguerias, E., and Denis, V. 2002. Biological variation of Loligo vulgaris (Cephalopoda: Loliginidae) in the eastern Atlantic and Mediterranean. Bulletin of Marine Science **71**(1): 515–534.
Murray, L.G., Hinz, H., Hold, N., and Kaiser, M.J. 2013. The effectiveness of using CPUE data derived from Vessel Monitoring Systems and fisheries logbooks to estimate scallop biomass. ICES Journal of Marine Science **70**(7): 1330–1340.
Nielsen, J.R. 2015. Methods for integrated use of fisheries research survey information in understanding marine fish population ecology and better management advice: improving methods for evaluation of research survey information under consideration of survey fish detection and catch efficiency. Wageningen University.
Nielsen, J.R., Thunberg, E., Holland, D.S., Schmidt, J.O., Fulton, E.A., Bastardie, F., Punt, A.E., Allen, I., Bartelings, H., and Bertignac, M. 2018. Integrated ecological–economic fisheries models—Evaluation, review and challenges for implementation. Fish and Fisheries **19**(1): 1–29.
Ord, J.K., and Getis, A. 1995. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geographical Analysis **27**(4): 286–306. doi:10.1111/j.1538-4632.1995.tb00912.x.
Pedersen, S.A., Fock, H.O., and Sell, A.F. 2009. Mapping fisheries in the German exclusive economic zone with special reference to offshore Natura 2000 sites. Marine Policy **33**(4): 571–590. doi:10.1016/j.marpol.2008.12.007.

688 Pennino, M.G., Conesa, D., Lopez-Quilez, A., Munoz, F., Fernández, A., and Bellido, J.M. 2016.
689 Fishery-dependent and-independent data lead to consistent estimations of essential
690 habitats. ICES Journal of Marine Science **73**(9): 2302–2310. Oxford University Press.
691 Pennino, M.G., Paradinas, I., Illian, J.B., Muñoz, F., Bellido, J.M., López-Quílez, A., and Conesa,
692 D. 2019. Accounting for preferential sampling in species distribution models.
693 Ecology and evolution **9**(1): 653–663.
694 Petitgas, P. 1997. Sole egg distributions in space and time characterised by a geostatistical
695 model and its estimation variance. ICES Journal of Marine Science **54**(2): 213–225.
696 Pinto, C., Travers-Trolet, M., Macdonald, J.I., Rivot, E., and Vermard, Y. 2019. Combining
697 multiple data sets to unravel the spatiotemporal dynamics of a data-limited fish
698 stock. Canadian Journal of Fisheries and Aquatic Sciences **76**(8): 1338–1349. NRC
699 Research Press.
700 Regimbart, A., Guitton, J., and Le Pape, O. 2018. Zones fonctionnelles pour les ressources
701 halieutiques dans les eaux sous souveraineté française. Deuxième partie : Inventaire.
702 Rapport d'étude. Les publications du Pôle halieutique A. Pôle halieutique
703 AGROCAMPUS OUEST, Rennes.
704 Rufener, M.-C., Kristensen, K., Nielsen, J.R., and Bastardie, F. 2021. Bridging the gap between
705 commercial fisheries and survey data to model the spatiotemporal dynamics of
706 marine species. Ecological Applications: e02453.
707 Salas, S., and Gaertner, D. 2004. The behavioural dynamics of fishers: management
708 implications. Fish and Fisheries **5**(2): 153–167. doi:https://doi.org/10.1111/j.1467-
709 2979.2004.00146.x.
710 Silvano, R.A., MacCord, P.F., Lima, R.V., and Begossi, A. 2006. When does this fish spawn?
711 Fishermen's local knowledge of migration and reproduction of Brazilian coastal
712 fishes. Environmental Biology of fishes **76**(2): 371–386.
713 Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E., and Semmens, B.X. 2019. The utility of
714 spatial model-based estimators of unobserved bycatch. ICES Journal of Marine
715 Science **76**(1): 255–267. doi:10.1093/icesjms/fsy153.
716 Thorson, J.T. 2018. Three problems with the conventional delta-model for biomass
717 sampling data, and a computationally efficient alternative. Canadian Journal of
718 Fisheries and Aquatic Sciences **75**(9): 1369–1382. NRC Research Press.
719 Thorson, J.T., Adams, C.F., Brooks, E.N., Eisner, L.B., Kimmel, D.G., Legault, C.M., Rogers, L.A.,
720 and Yasumiishi, E.M. 2020a. Seasonal and interannual variation in spatio-temporal
721 models for index standardization and phenology studies. ICES Journal of Marine
722 Science **77**(5): 1879–1892.
723 Thorson, J.T., Ciannelli, L., and Litzow, M.A. 2020b. Defining indices of ecosystem variability
724 using biological samples of fish communities: A generalization of empirical
725 orthogonal functions. Progress in Oceanography **181**: 102244.
726 doi:10.1016/j.pocean.2019.102244.
727 Tidd, A.N., Vermard, Y., Marchal, P., Pinnegar, J., Blanchard, J.L., and Milner-Gulland, E.J.
728 2015. Fishing for Space: Fine-Scale Multi-Sector Maritime Activities Influence Fisher
729 Location Choice. PLOS ONE **10**(1): e0116335. doi:10.1371/journal.pone.0116335.
730 Vermard, Y., Marchal, P., Mahévas, S., and Thébaud, O. 2008. A dynamic model of the Bay of
731 Biscay pelagic fleet simulating fishing trip choice: the response to the closure of the
732 European anchovy (Engraulis encrasicolus) fishery in 2005. Can. J. Fish. Aquat. Sci.
733 **65**(11): 2444–2453. doi:10.1139/F08-147.

734 Wikle, C.K., Zammit-Mangion, A., and Cressie, N. 2019. Spatio-temporal Statistics with R.
735 CRC Press.
736 Yan, Y., Cantoni, E., Field, C., Treble, M., and Flemming, J.M. 2022. Spatiotemporal modeling
737 of bycatch data: methods and a practical guide through a case study in a Canadian
738 Arctic fishery. Canadian Journal of Fisheries and Aquatic Sciences **79**(1): 148–158.
739 Yochum, N., Starr, R.M., and Wendt, D.E. 2011. Utilizing fishermen knowledge and expertise:
740 keys to success for collaborative fisheries research. Fisheries **36**(12): 593–605.
741

TABLES

Table 1. Ratio between the negative log-likelihood values (either commercial or scientific) from the IM accounting for PS and the IM ignoring PS.

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

Note. The ratio between negative log-likelihoods ($-\boldsymbol{log}(\boldsymbol{lkl})$) is given as: $\boldsymbol{r} = \frac{-log(lkl_{PS})}{-log(lkl_{noPS})}$.
If $\boldsymbol{r} < \mathbf{1}$, the model accounting for PS better fits the data than the model ignoring PS (no PS).

FIGURES CAPTION

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

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

Figure 3. Comparison between the observed scientific CPUE (y-axis) and the corresponding model predictions (x-axis) on the month of the survey, based on model integrating data from one commercial fleet only (either OTB_CEP, OTB_DEF, OTT_DEF) or from all commercial fleets (Integrated model). x-axis: model predictions. y-axis: scientific data observations (CPUE in kg/hour). Black line: linear regression 'log(scientific observations) ~ log(model predictions)'. r: Spearman correlation coefficient. Scientific data are integrated to inference for all models.

Figure 4. Sole case study. Comparison between predictions from the integrated model (using all fishing fleets) and the model integrating only one commercial fleet for the 12 months of year 2018. Left: OTB_CEP fleet, middle: OTB_DEF fleet, right: OTT_DEF fleet. x-axis: integrated model predictions. y-axis: single-fleet model predictions. The prediction values are log-scaled. Red points: predictions within the sampling area of the related fleets (i.e. the cells sampled by the fleet). Black points: predictions outside the sampling area of the related fleets. Black line: $x = y$ axis. Note that the intercept of the x-y line has been scaled to account for differences in the intercept values between models. Scientific data are integrated to inference for all models.

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

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

Figure 7. Left: index of persistence during the reproduction period of sole (February), whiting (March-May) and squids (January-April). Reproduction period defined from

789 ecological expertise. Index were computed from 2010 to 2018. Right: literature information
790 on reproduction grounds when available. For sole, the map represents egg concentration
791 from an egg and larvae survey conducted in 1982 (Arbault et al., 1986). For whiting, the
792 map represents records of age-2+ whiting (i.e. mature individuals), from two spring trawl
793 surveys that occurred between 1987 and 1992 (Houise and Forest, 1993). Model
794 predictions come from the integrated model accounting for PS.

795

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

799

800

FIGURES

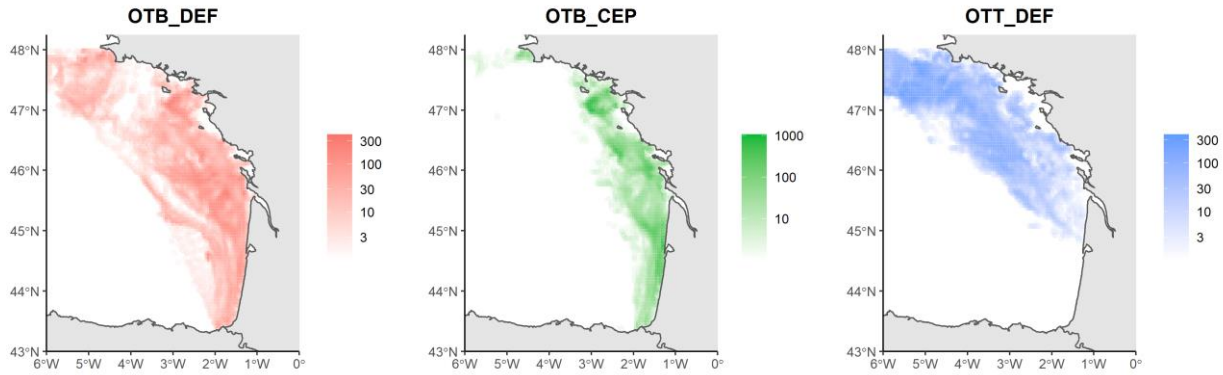


Figure 1

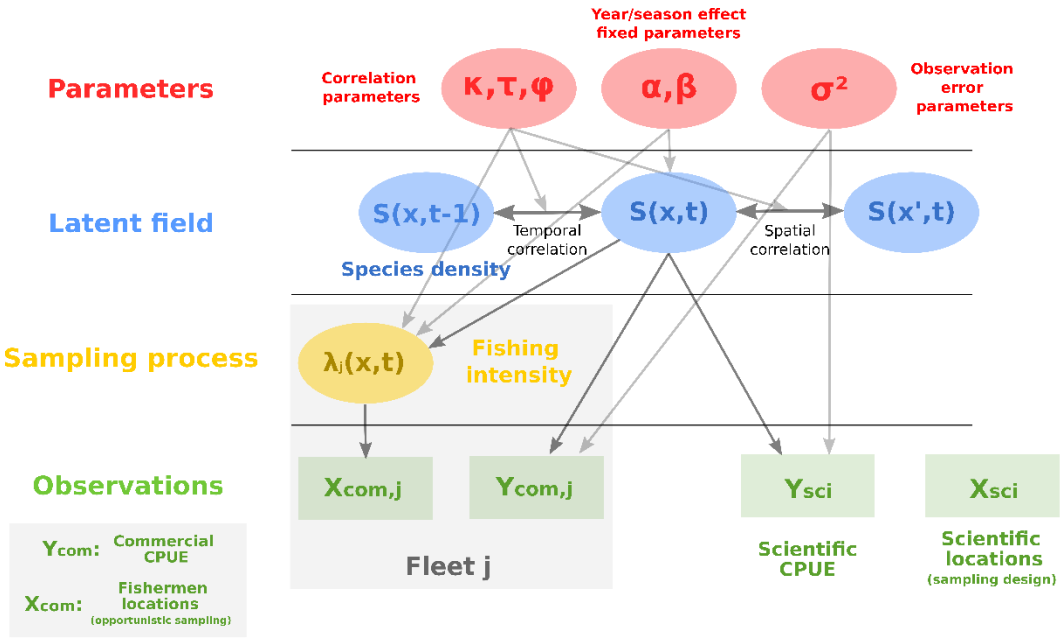


Figure 2

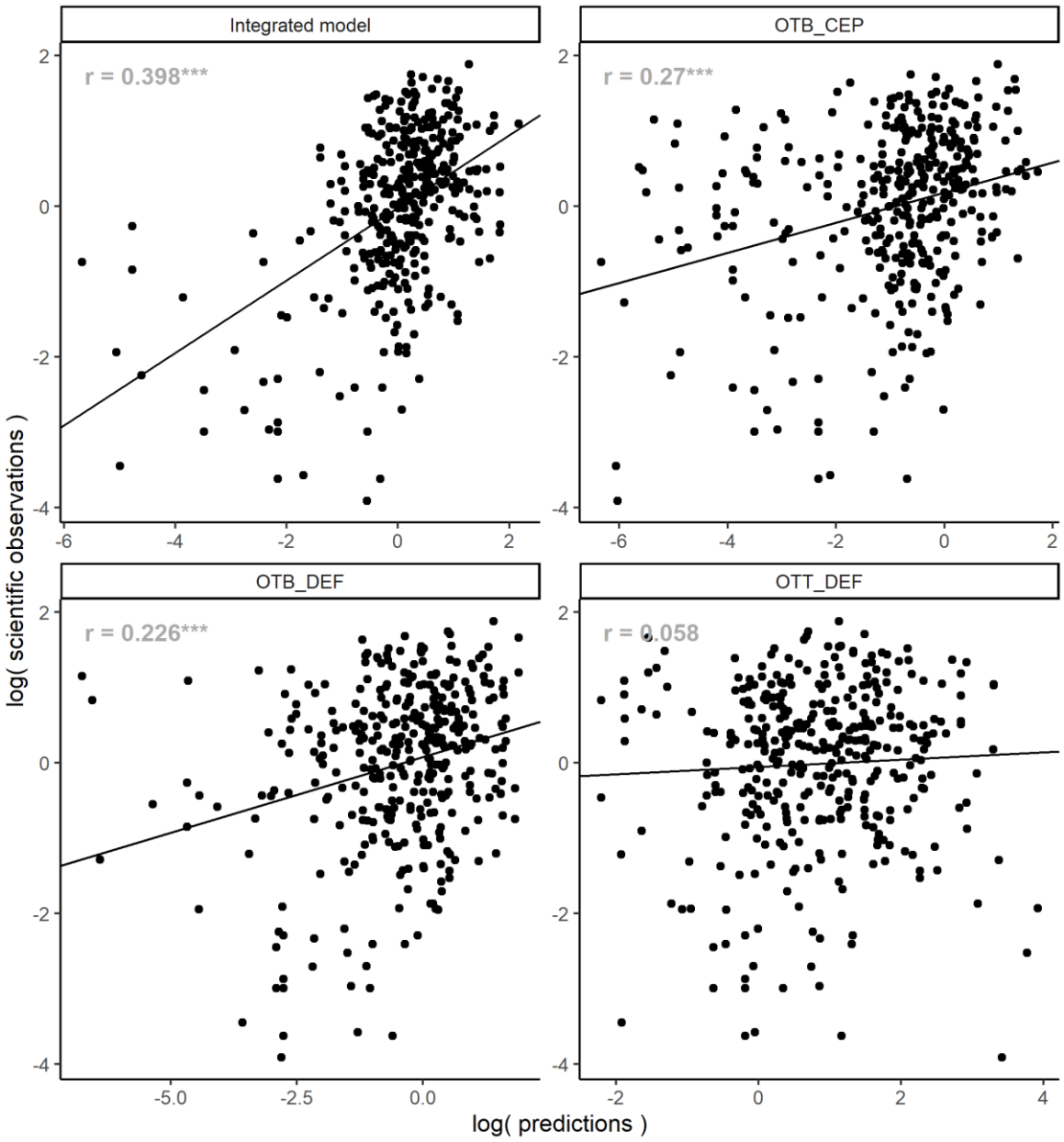

Integrated model
r = 0.398***
OTB_CEP
r = 0.27***
OTB_DEF
r = 0.226***
OTT_DEF
r = 0.058
log(scientific observations)
log(predictions)

Figure 3

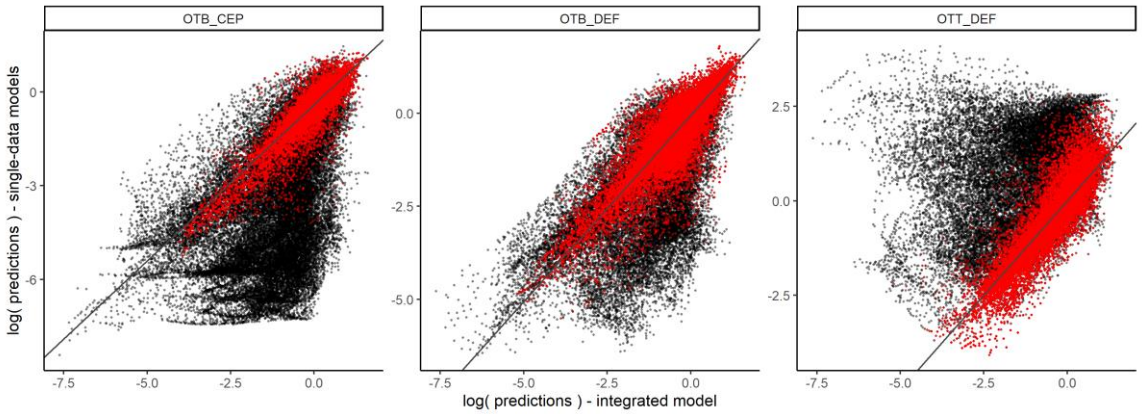
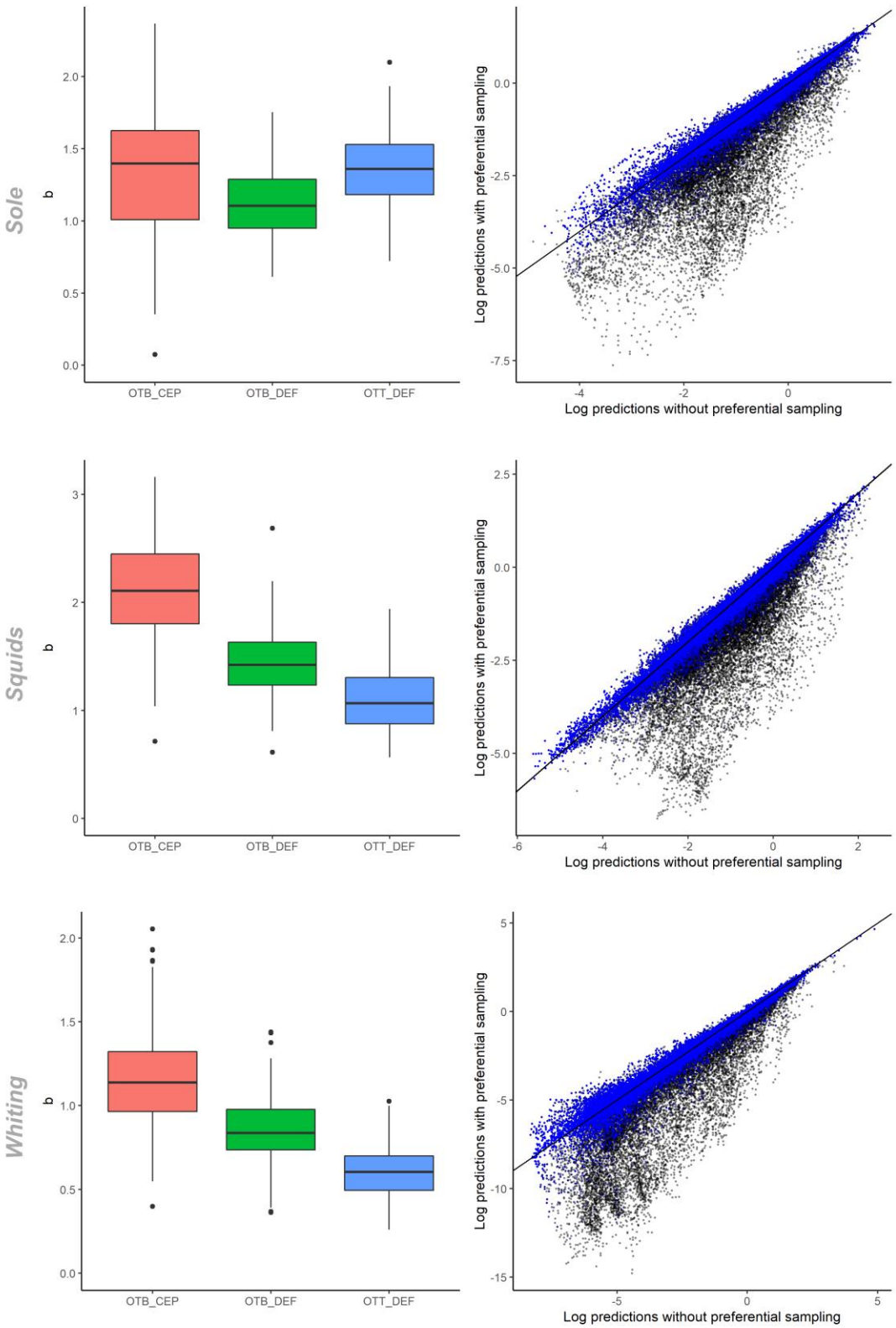

OTB_CEP
OTB_DEF
OTT_DEF
log(predictions) - single-data models
log(predictions) - integrated model
0
-3
-6
0.0
-2.5
-5.0
2.5
0.0
-2.5
-7.5
-5.0
-2.5
0.0

Figure 4



816

817 Figure 5

818

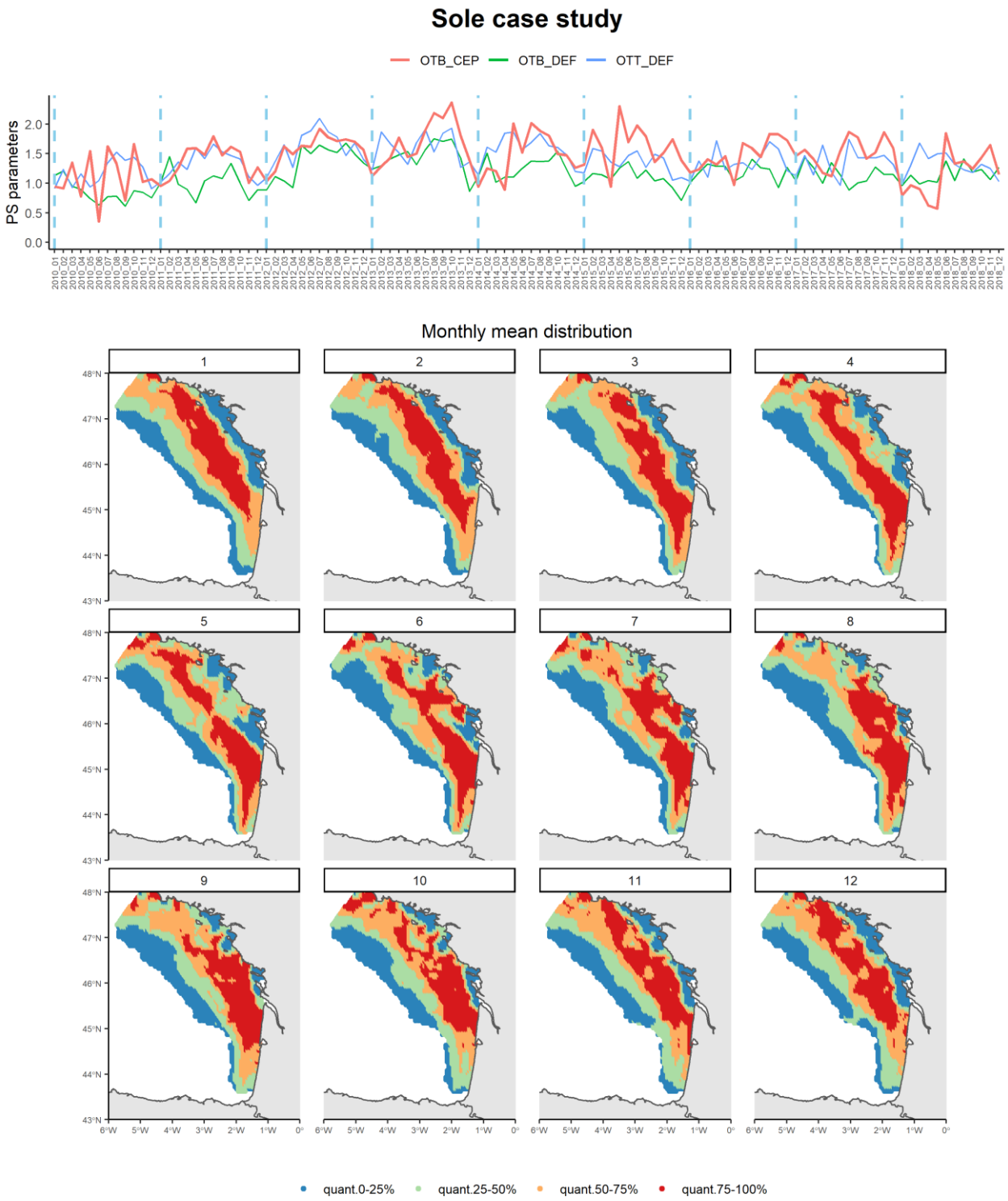
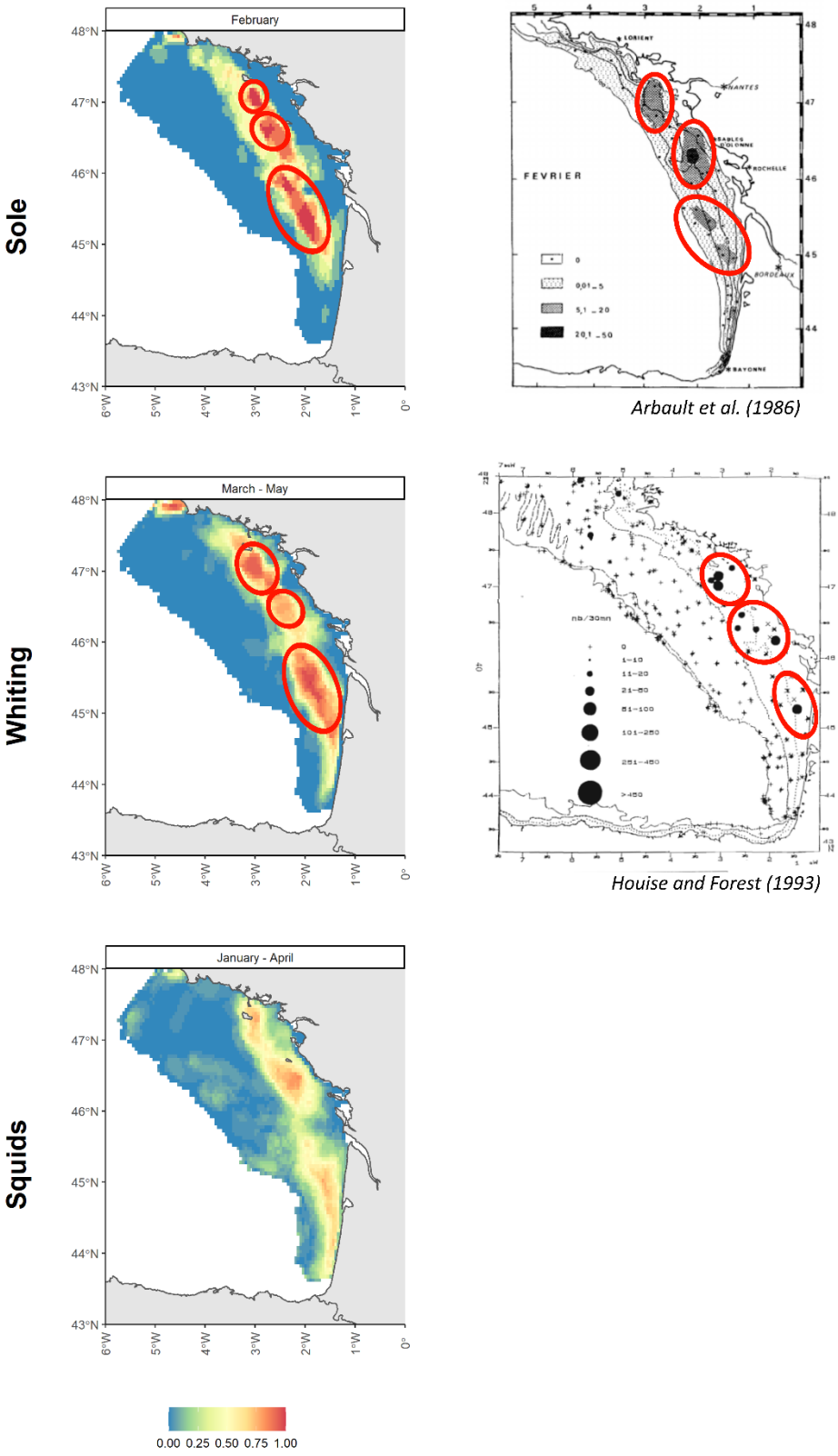

Sole case study
OTB_CEP
OTB_DEF
OTT_DEF
PS parameters
Monthly mean distribution
quant.0-25%
quant.25-50%
quant.50-75%
quant.75-100%

Figure 6



822

823 Figure 7

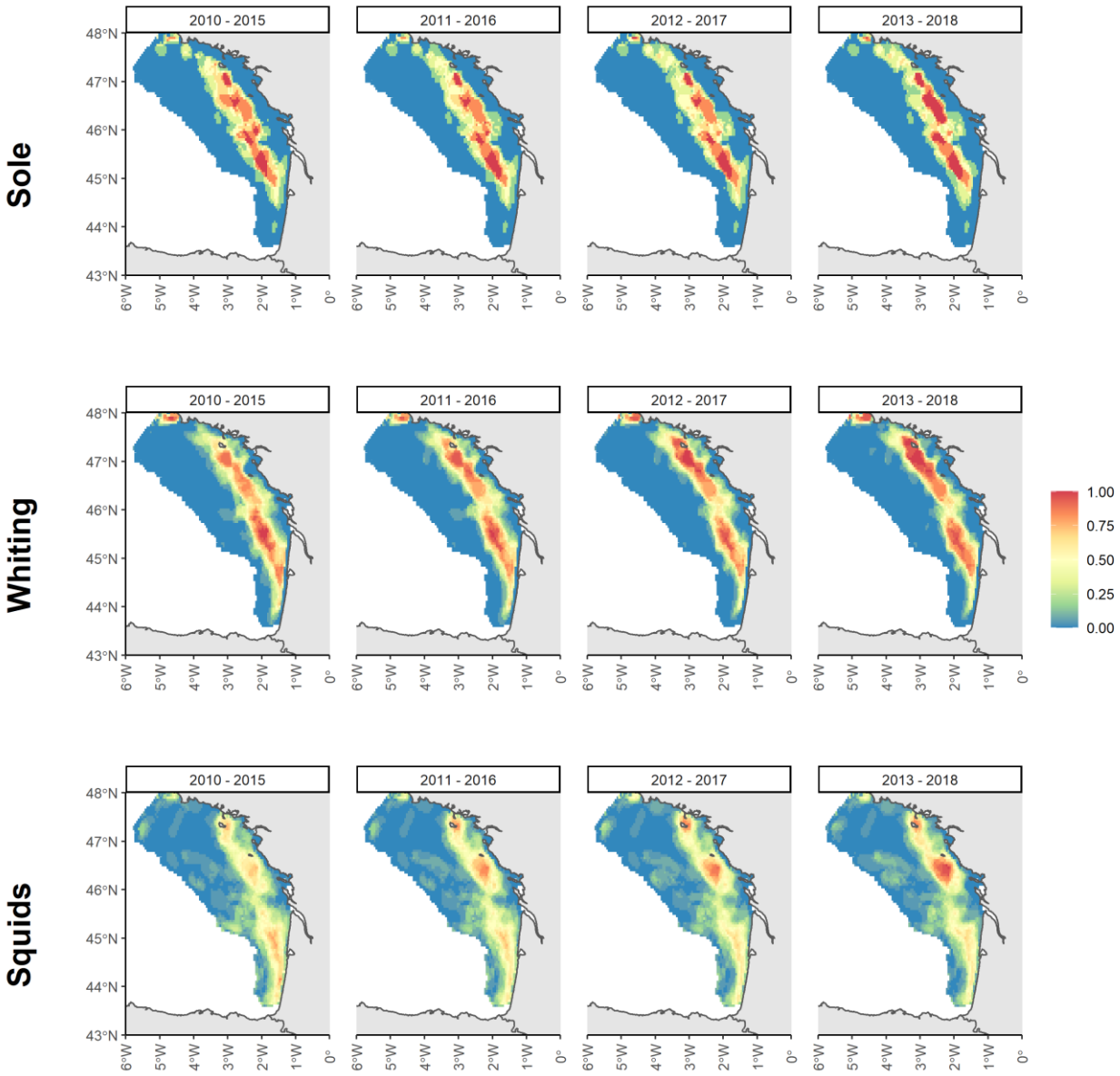


Figure 8

APPENDICES

The appendices are in the supplementary material file.