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

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

Canadian Journal of Fisheries and Aquatic Sciences
May 2023, Volume 80, Issue 5, Pages 808-824
https://doi.org/10.1139/cjfas-2022-0110
https://archimer.ifremer.fr/doc/00817/92910/

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

Alglave Baptiste [1, 2, *], Vermard Youen [1], Rivot Etienne [2], Etienne Marie-Pierre [3], Woillez Mathieu [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 : Baptiste Alglave, email address : baptiste.alglave@agrocampus-ouest.fr

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 fishermen 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.

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 standardized catchability and occur each year at the same period. Consequently, they provide standardized data on a large spatial extent for most species and size classes (Hilborn and Walters 2013; Nielsen 2015). 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. However, both scientific survey and onboard observer data 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 samples 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. However, 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 sufficient 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 landing per unit effort (LPUE) 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 fisher sampling behavior. Fisher typically tend to preferentially sample areas of higher biomass (a process referred to as preferential sampling, PS - Diggle et al. 2010). Hence, because data preferentially represent areas of highest biomass, ignoring PS in the distribution of fishing effort when estimating spatial distribution on larger areas can lead to overestimated biomass predictions (Conn et al. 2017; Pennino et al. 2019; Alglave et al. 2022).

In a recent paper, Alglave et al. (2022) developed an integrated modeling framework to infer spatial distribution of fish abundance by combining scientific survey CPUE and commercial LPUE data 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 this paper, we extend the modeling framework from Alglave et al. (2022) by adding 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 for distinct fishing fleets. To demonstrate 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 allows to combine the information from several fleets in order to produce accurate maps of spatio-temporal biomass. To illustrate the capacity of the framework to identify areas of aggregation during the spawning season, we processed model outputs to identify areas of recurrent aggregation occurring during the reproduction season and compared these to the information available in the literature.

2 MATERIAL AND METHODS

In this section, we first present the different species, the datasets and how we process and combine them to produce LPUE 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. In our applications, the models were fitted to data from 2010 to 2018 on a monthly time step (108 time steps). Then, we illustrate how the PS component modifies model predictions and can be interpreted, and how integrating several fleets in the analysis further improves model predictions. Last, we detail the method used to investigate spatio-temporal dynamics from model outputs and identify reproduction grounds based on the aggregation patterns of each of three species.

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). Reproduction period fall between January and April, but the peak of the reproduction fall in February. Discard rate is also very low, which makes the landings data a good proxy of the catch (ICES 2019a).

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

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

Overall scientific survey suggests that the predominant species in the Bay of Biscay is

Loligo Vulgaris (ICES, 2020a, p.17). All 3 species are data-poor: no information exists regarding their reproduction grounds but some information of the reproduction period exists for *Loligo Vulgaris* (Moreno et al. 2002). For this species, the reproduction period fall between January and April.

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 LPUE 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, we used the term Landing Per Unit of Effort (LPUE) to refer to commercial observations expressed in kg of (mature) biomass per hour fished. Discards are neglected in our approach, hence LPUE are considered as biomass indices.

Our model can integrate data of different fishing fleets. For the purpose of our application, we selected three different metiers that belong to the same fleet of trawlers (in this case, the 'métier' term refers to a combination of gear and of a set of species that are targeted by the vessels – see the Data Collection Framework and EC (2008)): 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). In the following those three métiers are referred as three fleets. These three fleets were selected as they offer three main advantages. First, their targeting behaviors and technical characteristics are similar. Second, catch per unit effort of trawlers are generally good indicator of fish relative abundance while other gears (longline, gillnet) may face saturation effects leading to non-linear relationship between catches and fishing time (Hovgêrd and

Lassen 2008). Third, the combination of the three fleets cover the full spatial domain (Figure 1).

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

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

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

(2) to reallocate the logbooks declaration on the related VMS fishing locations. This reallocation is realized individually for each fishing vessel trajectory by uniformly reallocating the landings on all fishing locations. The link between both data sources is realized through the combination ‘vessel identifier x statistical

rectangle x fishing trip x day'. LPUE are then computed by simply dividing the reallocated landings by the related fishing time.

2.2.2 Scientific data

We also integrated scientific data in the analysis. For whiting and squids we used the survey data from the EVHOE survey. The Orhago survey was used for sole (ICES, 2020 - see SM3, 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.

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

2.3 Spatio-temporal integrated model

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

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

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

This first model was purely spatial as no temporal dimension was included in the model. In this paper, we extend the framework by incorporating a temporal component to model the evolution of the latent field of biomass across the monthly time steps (Figure 2).

2.3.1 Biomass field

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

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

where $x \in D \subset R^2$ stands for the spatial locations and $t \in [\![1,T]\!]$ for the monthly time steps. $S(x,t)$ is in the same unit as the data (here kg/hr fished as data are CPUE for survey trawls or LPUE for commercial landing declarations). The term $\alpha_S(t)$ is a time varying intercept modeled 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 all time steps, the temporal correlation is a critical component that 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):

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

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

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

2.3.2 Sampling process for the commercial fishing points

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

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

$$log(\lambda_j(x,t)) = \alpha_{Xj}(t) + b_j(t) \cdot log(S(x,t)) + \eta_j(x,t) \qquad (3)$$

where:

- the time varying intercept $\alpha_{Xj}(t)$ quantifies the average fishing intensity on the whole area; it is modeled as a fixed effect;
- the time varying $b_j(t)$ quantifies the strength of PS towards biomass $S(x,t)$; it is modeled as a fixed effect too. If $b_j(t) = 0$, then PS is null. If $b_j(t) > 0$, then PS occurs and the greater, the stronger PS. Alternatively, $b_j(t) < 0$ would indicate that fisher have a repulsive behavior towards the resource.
- 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. 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). The observation model explicitly considers that each fleet can have its own catchability and its own zero inflation parameter.

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

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

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

$\mu_j(x_i,t_i) = q_j.S(x_i,t_i)$ is the expected catch of fleet j at location x_i and time step t_i. It is the product of the latent field value $S(x_i,t_i)$ and of the relative catchability coefficient specific for fleet j denoted q_j. Specifically for each fleet j, ξ_j is a zero-inflation parameter controlling the proportion of zero in the data, σ_j^2 is the observation variance when the catch is positive.

Equation (4) shows the two components that compose the probability to observe a catch y_i:

- the probability to obtain a zero catch ($y_i = 0$). It is modeled as a Bernoulli variable with probability $p_i = \exp\left(-e^{\xi_j} \cdot \mu_j(x_i,t_i)\right)$. p_i is equivalent to the probability to obtain a 0 value with a Poisson distribution of intensity $e^{\xi_j} \cdot \mu_j(x_i,t_i)$. The value of ξ_j controls the intensity of the zero inflation, when ξ_j increases, the amount of zero in the data decreases. Then the probability to obtain a positive catch is given by $1 - p_i$.

- the value of the positive catch is modeled through a lognormal distribution L with expected value $\frac{\mu_j(x_i,t_i)}{(1-p_i)}$ and observation error σ_j^2. The standardization by $(1-p_i)$ allows to keep the expectancy of the observation model to $\mu_j(x)$.

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

$$q_j = k_j.q_{ref} \qquad (6)$$

2.3.4 Maximum likelihood estimation

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

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 on a single time step (here November 2018).

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

2.5 Evaluating the value of modeling PS

2.5.1 Comparing the inferences with and without 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.

2.5.2 Interpreting the intensity of preferential sampling

The estimates of PS parameters $b_j(t)$ in eq. (3) 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 among the three species and the different fleets. Then, focusing on

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

315 **2.6 Investigating spatio-temporal distribution and identifying reproduction**
316 **grounds**

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

320 ***2.6.1 Aggregation index***

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

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

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

344 ***2.6.2 Comparing the results with the available literature***

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

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

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

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

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

368 To assess whether the aggregation patterns within the reproduction period are stable over
369 the time period, we iteratively computed the persistence index over a 5-year mobile time-
370 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 predictions with survey data (for the month of the survey), integrating several fleets into the analysis improves correlation with scientific data (Figure 3). It also reduces the standard deviation of the predictions on the full domain (SM 7, Figure S7).

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, when using single fleet data, predictions realized outside the spatial range of the fleet largely depart from the ones realized through the integrated models (black dots, Figure 4), emphasizing the contribution of the other fleets to improve inferences on areas poorly covered by single fleet. 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 in eq. (3)) for the different species, the different fleets and the different time steps provide information on the targeting behavior that 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 ($b_{OTB_{CEP}} > b_{OTB} > b_{OTT}$); this is consistent with species

spatial distribution as whiting (like squids) are found in coastal areas where the OTB_CEP fleet is preferentially operating (Figure 1). 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 which is also consistent with expertise as those three fleet target this high commercial value species.

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 inferences

Because estimates of b are positive, spatial density of fishing points is positively correlated with biomass density. Then logically, ignoring PS leads to a positive bias (overestimate) in biomass estimates in areas not sampled by the commercial fleets compared to estimates obtained while considering (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 for both the likelihood associated with the commercial and the scientific data (Table 1). For whiting and squids, there are no strong modifications in both scientific and commercial likelihoods. Overall, accounting or not for PS does not strongly modify the overall pattern of species abundance (SM 8).

3.4 Investigating spatio-temporal dynamics of fish biomass

Results provide biomass density maps on a monthly time step that emphasize 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 between months temporal correlations in the biomass field values. The range parameters are estimated to 55 km for sole and squids while being estimated to 67 km for whiting emphasizing wider spatial autocorrelation for this species.

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

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

3.5 Aggregation index and reproduction grounds

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

443 the eggs and larval drift is 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 whitings
447 (Houise 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 Houise 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.

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 complements 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 with regards to the joint dynamics of species distribution and fishing 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 contrast 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 that remain unsampled.

Filtering the mature fraction of the population in both the scientific and the commercial data make possible to infer the spatio-temporal distribution of the fraction of the biomass that could be potentially mature through the year on a monthly time step. We developed an index to infer aggregation areas of the potentially 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 enables 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).

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).

Our results could also be combined with fisher 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 fisher. 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.

Our model remains relatively simple with regards to all the temporal processes that actually occur within a fishery. It is both a strength and a weakness: 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. 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 yearly and seasonal terms in addition to some 'season x year' interaction terms as 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, even though it provides a nice conceptual view of seasonality, complexifying our model in that direction should not deeply modify our inference of the biomass field. 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 field of biomass (Cao et al. 2020). This would require to refine further the demographic resolution of the VMS-logbooks data (see for instance Azevedo and Silva 2020), but once done, it would give access to huge data for inferring the space-time dynamics of fish populations. Finally, our model considers fisher preferentially sample areas where the biomass is higher (preferential sampling), but does not consider any other drivers and specifically the temporal and spatio-temporal relations that can affect fishers behavior. These can be highly complex and may depend on the distribution of the resource, tradition/habits, management regulations (Abbott et al., 2015; Girardin et al., 2017; Salas and Gaertner, 2004; Hintzen, 2021). These drivers are rarely studied in both space and time (although see Tidd et al., 2015). Our framework could allow to jointly model the dynamics of the species, the distribution of the effort, the link that relates species distribution and effort in space/time and all the other spatial and/or temporal drivers that affect the distribution of fishing effort. For instance, we could relate the fishing intensity to the biomass field from the previous time steps, or alternatively consider that the locational choice depends on the catches of the previous time steps. Adding such covariates and spatio-temporal dependencies in the sampling equation (eq.3) will probably not modify the overall pattern of biomass distribution, but it would make possible to quantify the drivers 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

1993; Moreno et al. 2002). Several methods exist and could be adapted to extract the spatial patterns that shape model outputs, their related temporal variation and identify 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). While the approach we adopted in the manuscript requires to know the reproduction period of the species and would be inappropriate in a context of a changing reproduction time-span, those alternative methods would not require any a priori. Hence, these methods would be more appropriate to identify phenological modifications in species life cycle in response to climate change. Applying those kind of methods to the huge amount of data available from mandatory declarations, would make possible to track the effect of climate change on fish phenology at a monthly/seasonal scale, while it is generally only possible at a yearly time step through scientific survey data (Maureaud et al. 2020).

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

Future use for Marine Spatial Planning

Our model has potential application for Marine Spatial Planning (MSP). Janßen et al. (2018) highlighted that one of the main requirements for implementing MSP 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

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 Loiselle 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

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.

FUNDINGS

The authors declare no specific funding for this work.

COMPETING INTERESTS

The authors declare there are no competing interests.

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.

CODES

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

https://github.com/balglave/sdm_vms_logbooks

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.

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 francais - 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.

Maureaud, A., Frelat, R., Pécuchet, L., Shackell, N., Mérigot, B., Pinsky, M.L., Amador, K., Anderson, S.C., Arkhipkin, A., Auber, A., Barri, I., Bell, R.J., Belmaker, J., Beukhof, E., Camara, M.L., Guevara-Carrasco, R., Choi, J., Christensen, H.T., Conner, J., Cubillos, L.A., Diadhiou, H.D., Edelist, D., Emblemsvåg, M., Ernst, B., Fairweather, T.P., Fock, H.O., Friedland, K.D., Garcia, C.B., Gascuel, D., Gislason, H., Goren, M., Guitton, J., Jouffre, D., Hattab, T., Hidalgo, M., Kathena, J.N., Knuckey, I., Kidé, S.O., Koen-Alonso, M., Koopman, M., Kulik, V., León, J.P., Levitt-Barmats, Y., Lindegren, M., Llope, M., Massiot-Granier, F., Masski, H., McLean, M., Meissa, B., Mérillet, L., Mihneva, V., Nunoo, F.K.E., O'Driscoll, R., O'Leary, C.A., Petrova, E., Ramos, J.E., Refes, W., Román-Marcote, E., Siegstad, H., Sobrino, I., Sólmundsson, J., Sonin, O., Spies, I., Steingrund, P., Stephenson, F., Stern, N., Tserkova, F., Tserpes, G., Tzanatos, E., van Rijn, I., van Zwieten, P.A.M., Vasilakopoulos, P., Yepsen, D.V., Ziegler, P., and Thorson, J. 2020. Are we ready to track climate-driven shifts in marine species across international boundaries? - A global survey of scientific bottom trawl data. Global Change Biology: gcb.15404. doi:10.1111/gcb.15404.

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.

Pennino, M.G., Conesa, D., Lopez-Quilez, A., Munoz, F., Fernández, A., and Bellido, J.M. 2016. Fishery-dependent and-independent data lead to consistent estimations of essential habitats. ICES Journal of Marine Science **73**(9): 2302–2310. Oxford University Press.

Pennino, M.G., Paradinas, I., Illian, J.B., Muñoz, F., Bellido, J.M., López-Quílez, A., and Conesa, D. 2019. Accounting for preferential sampling in species distribution models. 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

TABLES

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

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

If $r < 1$, 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. Coordinate system: WGS84.

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

Figure 3. Sole case study. 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. *** stands for the level of significance. Each point is a grid cell in which fall a scientific data point. We compare the scientific observation values to the related prediction values.

Figure 4. Sole case study. Comparison between predictions (each point is a combination 'grid cell x time step'.) 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 all 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

936 represented. Model predictions come from the integrated model accounting for PS.
937 Coordinate system: WGS84.

938

939 Figure 7. Left: index of persistence (average over 2010-2018) during the reproduction
940 period of sole (February), whiting (March-May) and squids (January-April). Reproduction
941 period defined from ecological expertise. Right: literature information on reproduction
942 grounds when available. For sole, the map represents egg concentration from an egg and
943 larvae survey conducted in 1982 (Arbault et al., 1986). For whiting, the map represents
944 records of age-2+ whiting (i.e. mature individuals), from two spring trawl surveys that
945 occurred between 1987 and 1992 (Houise and Forest, 1993). Model predictions come
946 from the integrated model accounting for PS. Coordinate system: WGS84.

947

948 Figure 8. Persistence indices within the reproduction period computed on a 5-years
949 mobile time-span for each 3 species (5-years time span indicated on the top of each map).
950 Model predictions come from the integrated model accounting for PS. Coordinate system:
951 WGS84.

952

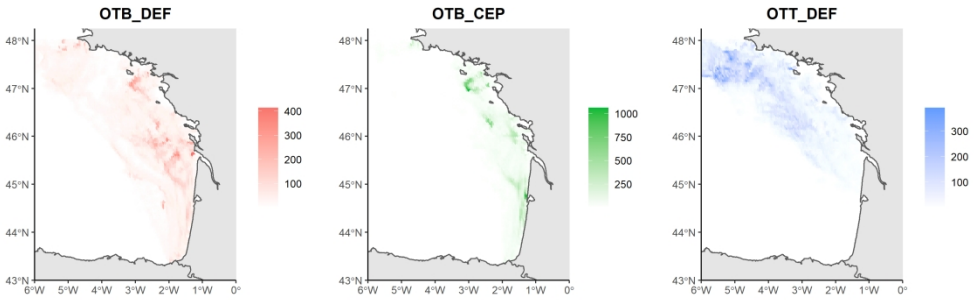

Figure 1

774x258mm (118 x 118 DPI)

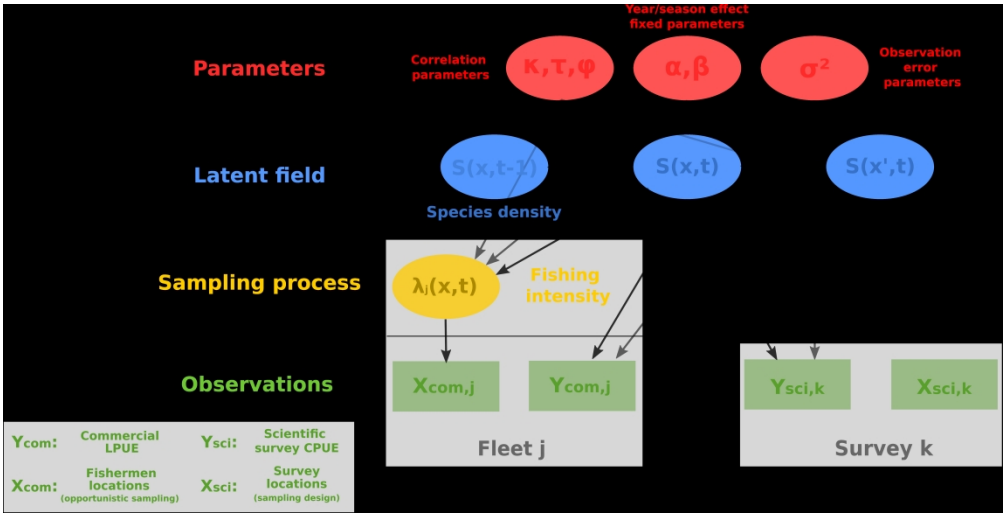

Figure 2

1054x535mm (197 x 197 DPI)

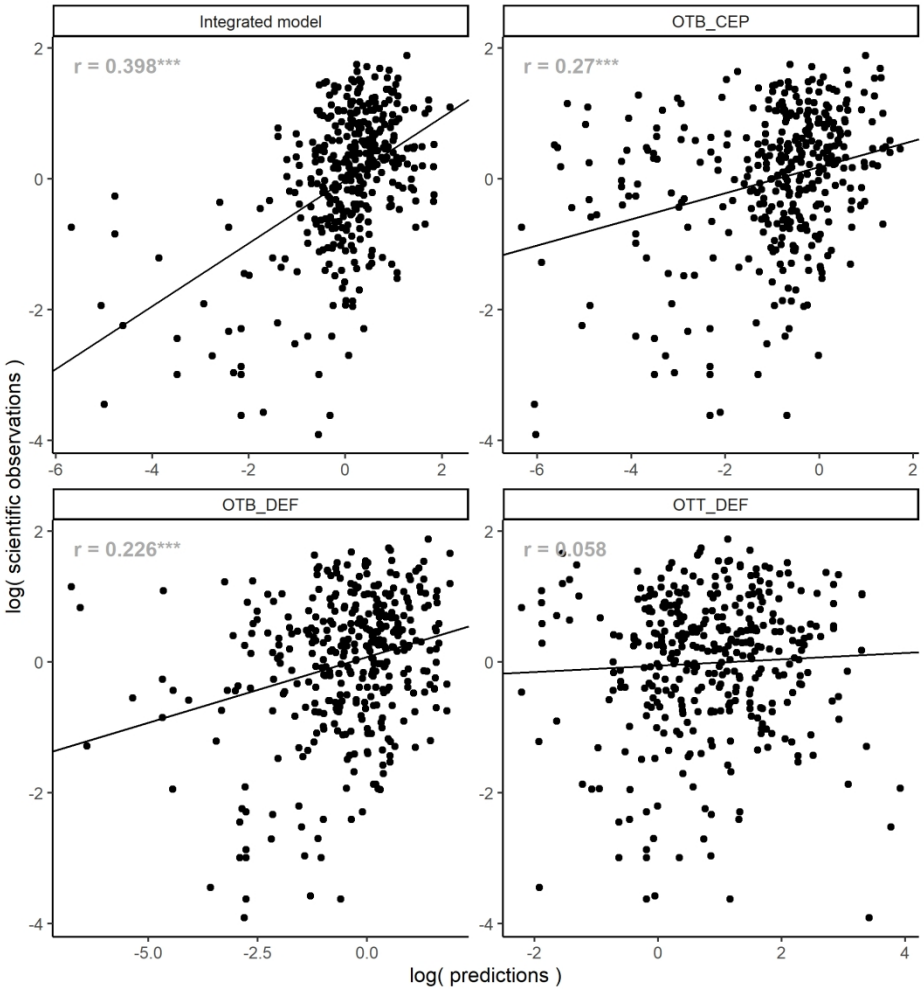

Figure 3

484x484mm (118 x 118 DPI)

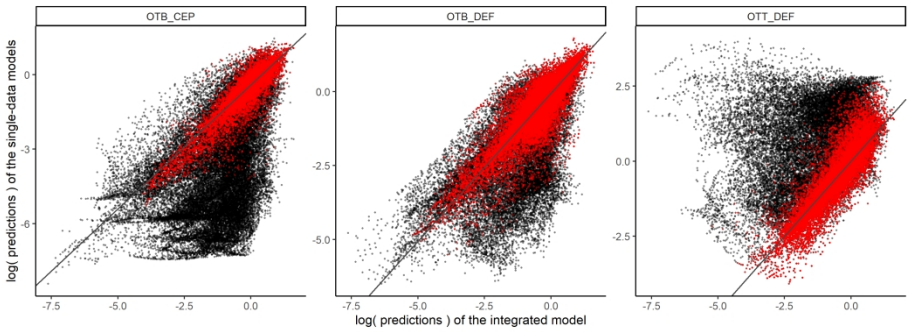

Figure 4

774x258mm (118 x 118 DPI)

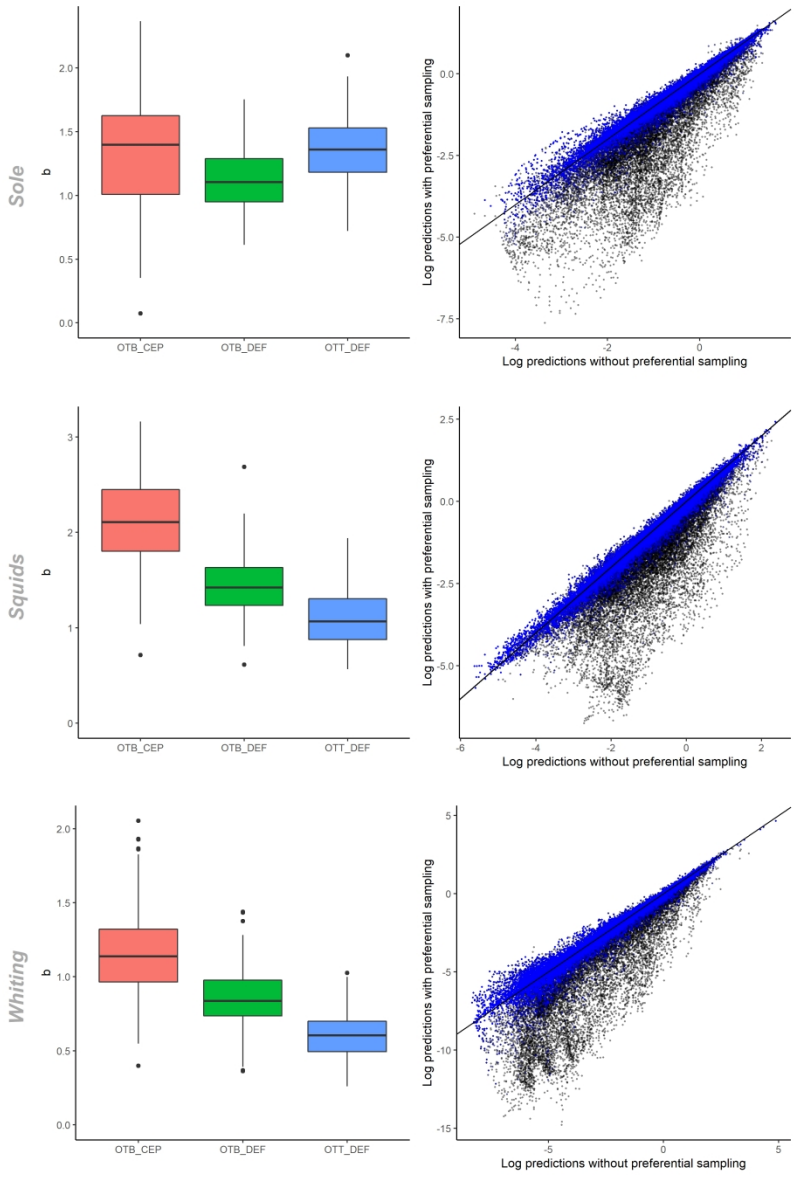

Figure 5

645x968mm (118 x 118 DPI)

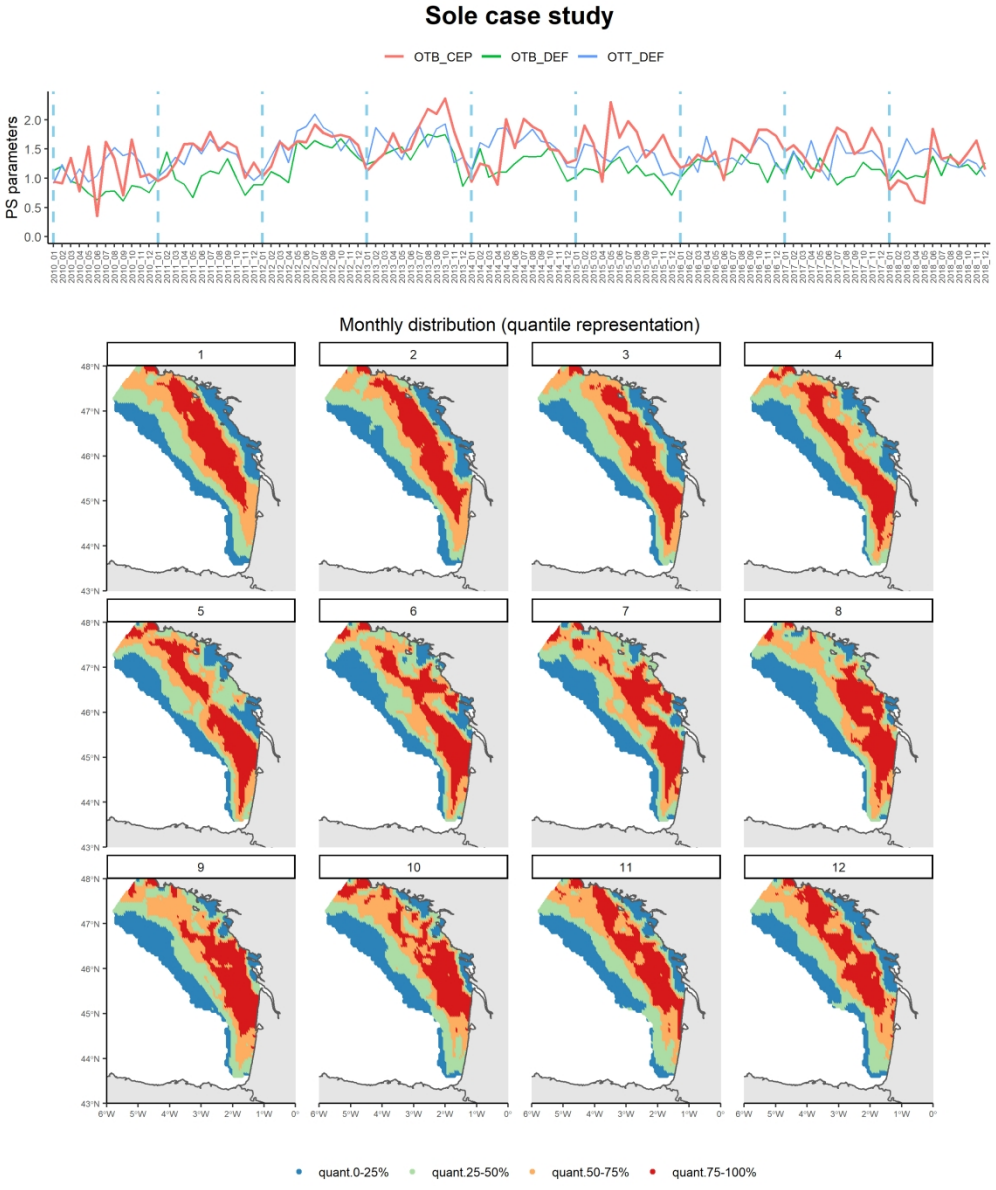

Figure 6

645x774mm (118 x 118 DPI)

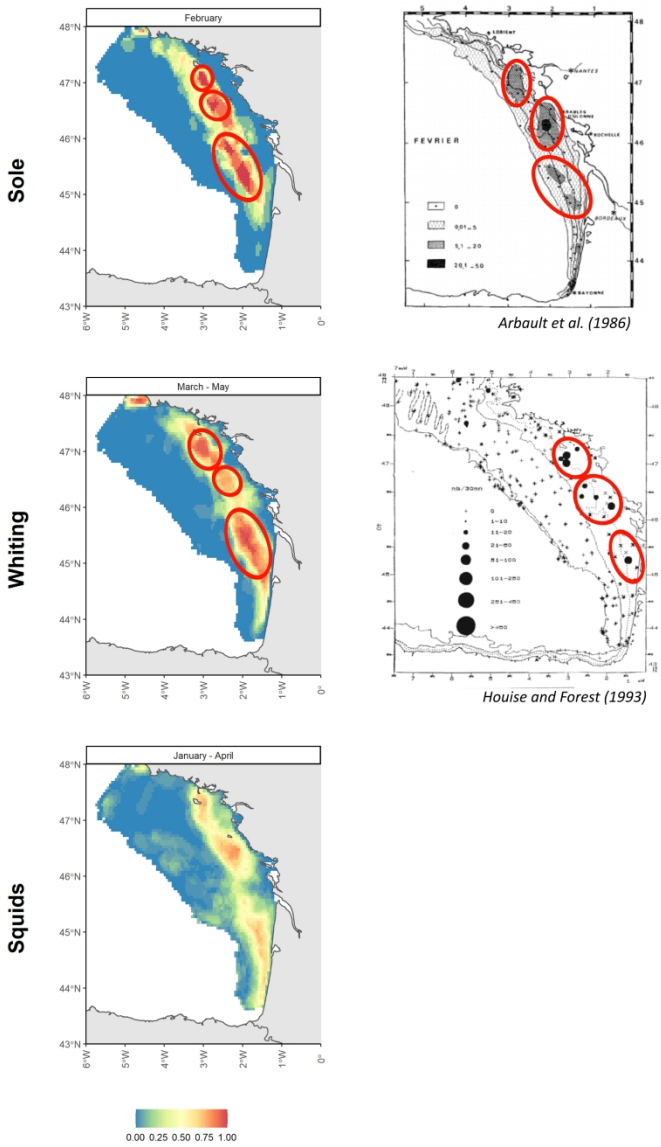

Figure 7

477x793mm (118 x 118 DPI)

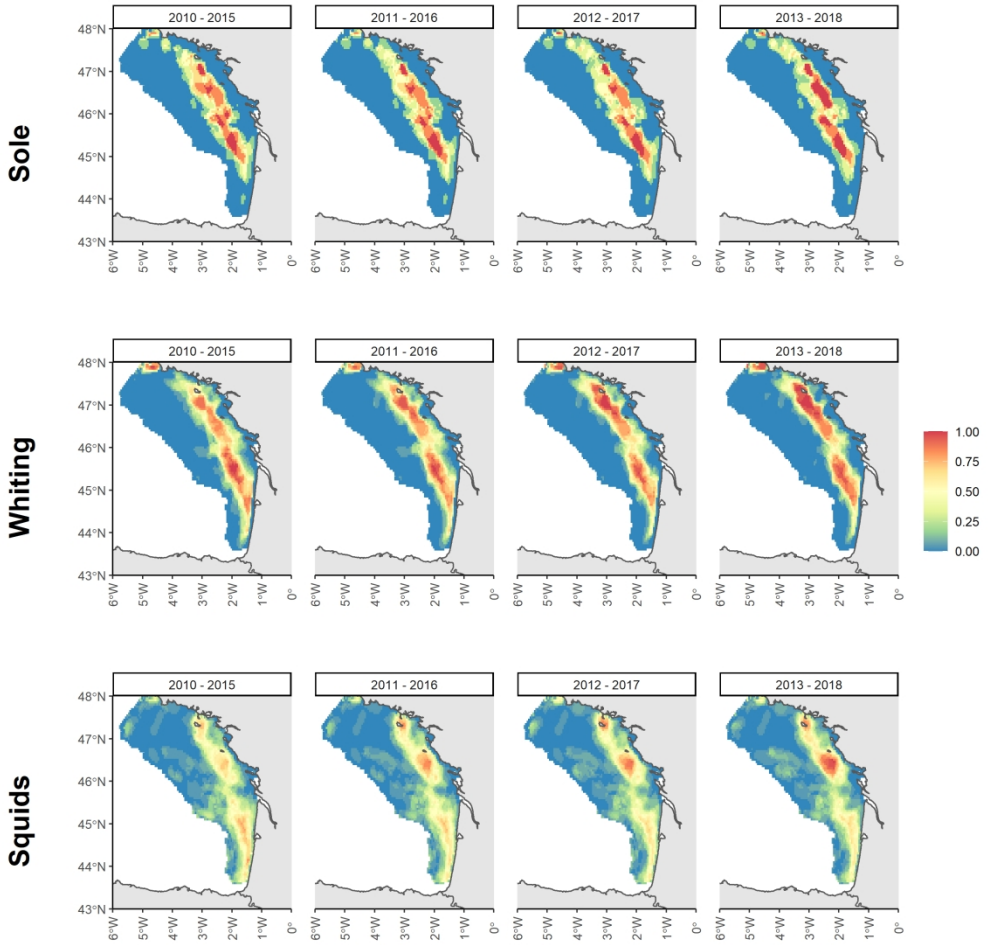

Figure 8

645x645mm (118 x 118 DPI)