

# Supplementary material

## 1 FILTERING THE MATURE FRACTION FROM THE LANDINGS

The filtering of the mature fraction of the landings is described through the formula:

$$Land_m(CC) = Land(CC) \cdot p_m(CC)$$

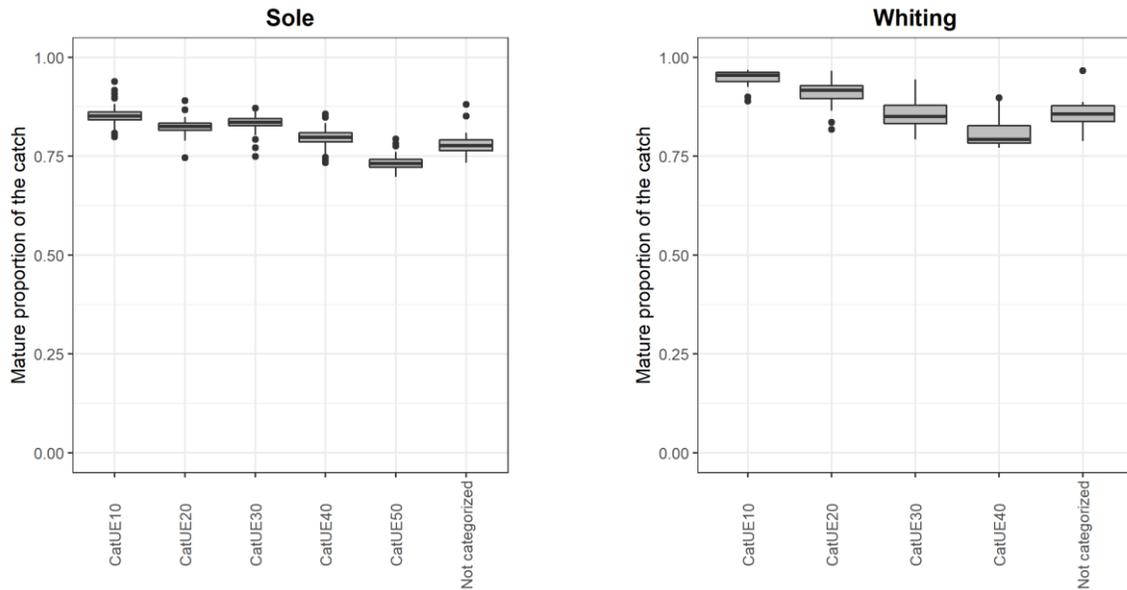
$$p_m(CC) = \sum_l p_l(CC, l) * MO(l)$$

where  $Land$  are the landings and  $Land_m$  the mature fraction of the landings.

By merging sales notes and logbooks, a comprehensive part of the landings can be expressed by commercial size categories  $CC$  so as landings can be written as  $Land(CC)$  (SACROIS - Demanèche et al., 2013)). These commercial categories are regularly sampled to derive length structure of each commercial category. These data are often used in stock assessment routines to obtain catch-at-length data (ICES, 2017). The proportion of length class  $l$  within commercial category  $CC$  is denoted  $p_l(CC, l)$ . To compute the mature proportion of the corresponding commercial category  $p_m(CC)$ ,  $p_l(CC, l)$  is combined with the proportion of mature individuals  $MO(l)$  for a specific length class  $l$  available through maturity ogive. Once the mature proportions per commercial category are available, they can be crossed with landings  $Land(CC)$  to obtain the mature fraction of the landings  $Land_m(CC)$ .

In the case studies, auction data are taken from ObsVentes data (Vigneau, 2009). The maturity ogive is assumed constant over the full period and the size distribution within each catch category is assumed to vary on a quarterly time step (sampling of demographic data is designed by quarters). Overall, the proportions of mature individuals within each commercial category fall between 75% to 100% for both sole and whiting (see SM, [Figure](#)

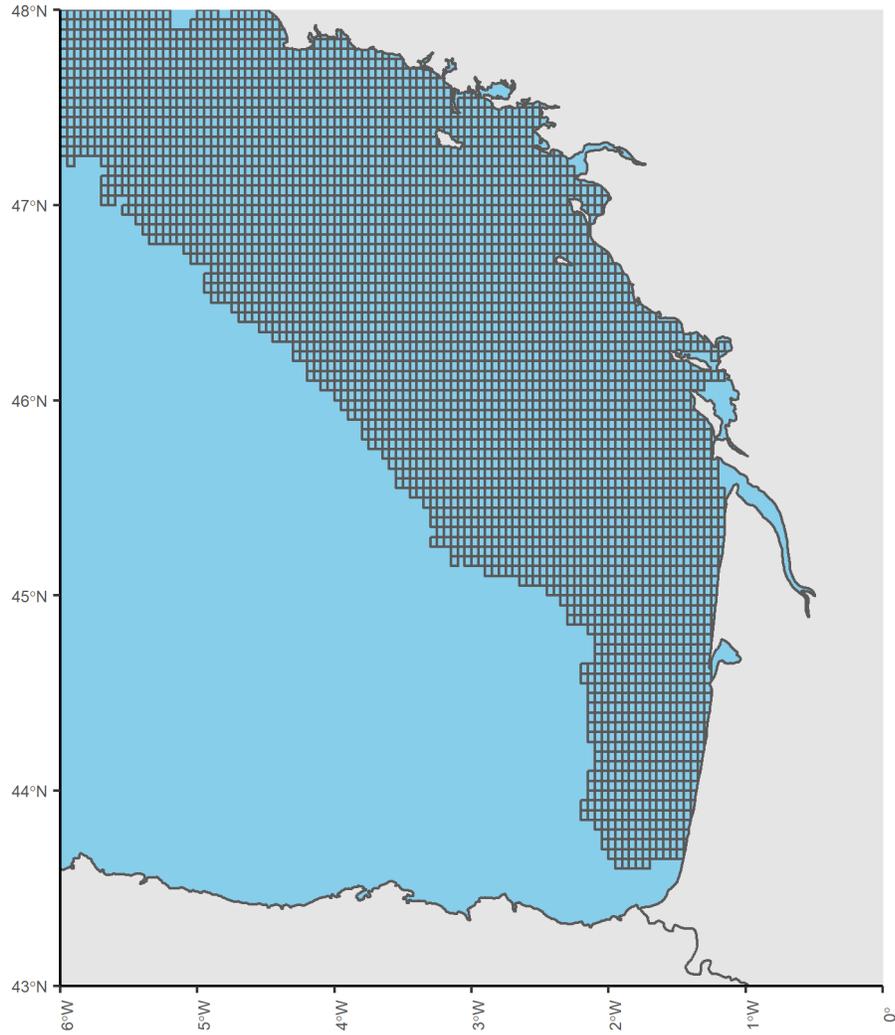
23 **S1**). When data are lacking for a specific quarter (because sampling is missing for certain  
24 quarters and commercial category), the missing data is replaced by the average of the  
25 mature proportion for the related catch category over the period.



26  
27 **Figure S1.** *Mature proportion of the catch per commercial size category on the whole*  
28 *time series (each point is the record of a quarter). Only the commercial categories where*  
29 *more than 10 individuals have been observed are plotted.*

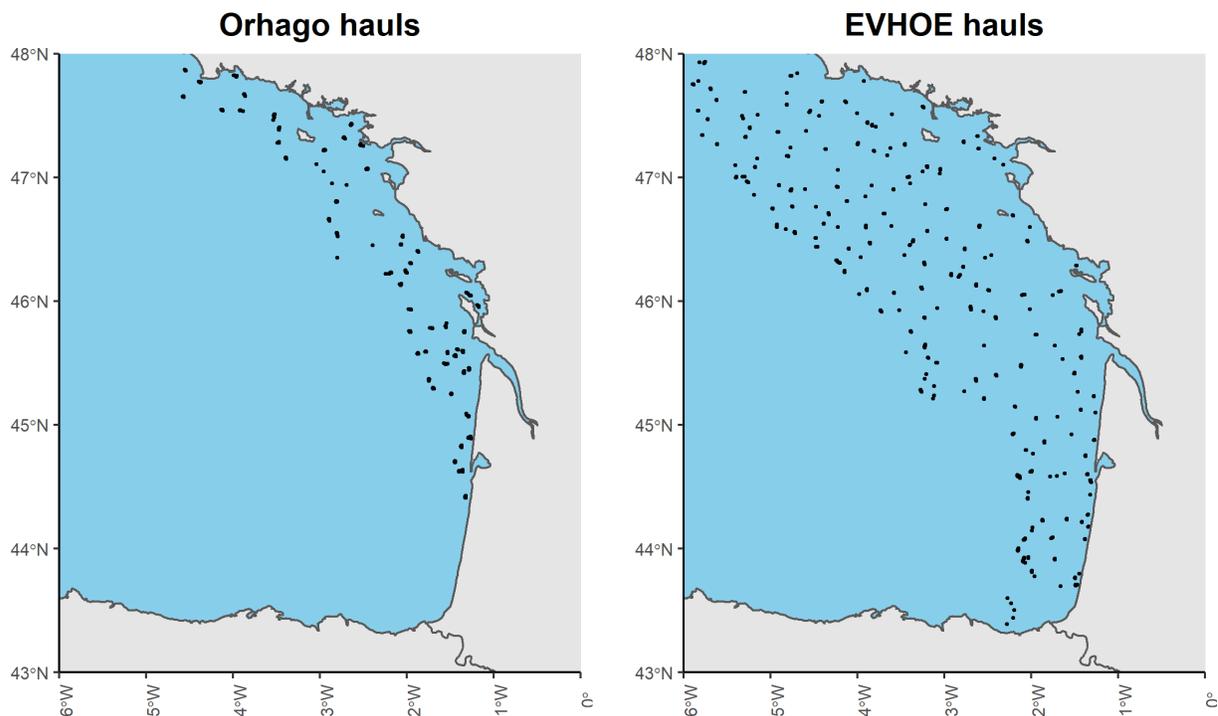
30

31 **2 DISCRETIZATION GRID**  
32



33  
34 **Figure S2.** Grid used to discretize commercial data, integrate the point process  
35 constant and compute the biomass predictions. Resolution: 0.05°x0.05°.  
36  
37  
38

39 **3 SURVEY INFORMATION AND SAMPLING LOCATIONS**  
40



41  
42 **Figure S3.** *Spatial distribution of the fishing hauls of the Orhago and the EVHOE*  
43 *surveys from 2010 to 2018.*

44  
45 Orhago is an annual beam trawl survey occurring in November and designed to assess  
46 sole stock status in the Bay of Biscay. Each year 50 stations are sampled within 4 strata  
47 all along the Bay of Biscay. Note that this survey is mainly coastal and does not sample  
48 offshore areas. EVHOE is an annual bottom trawl survey occurring in late October,  
49 November and early December. It is designed for demersal fishes in the Bay of Biscay  
50 and in the Celtic Sea. In the Bay of Biscay, 80 to 90 sampling hauls are recorded each  
51 year.

#### 52 4 ZERO-INFLATED OBSERVATION MODEL

53 The probability to obtain catch  $y_i$  conditionally on the location  $x_i$ , the time-step  $t_i$ , the  
54 biomass field value  $S(x_i, t_i)$  and the fleet  $j$  is expressed as follow:

$$55 \quad P(Y_i = y_i | x_i, t_i, S(x_i, t_i), j) = \begin{cases} p_i & \text{if } y_i = 0 \\ (1 - p_i) \cdot L\left(y_i, \frac{\mu_j(x_i, t_i)}{(1 - p_i)}, \sigma_j^2\right) & \text{if } y_i > 0 \end{cases}$$

$$56 \quad p_i = \exp\left(-e^{\xi_j} \cdot \mu_j(x_i, t_i)\right)$$

57  
58  $\mu_j(x_i, t_i) = q_j \cdot S(x_i, t_i)$  is the expected catch of fleet  $j$  at location  $x_i$  and time step  $t_i$ . It is  
59 the product of the latent field value  $S(x_i, t_i)$  and of the relative catchability coefficient of  
60 fleet  $j$  denoted  $q_j$ .  $\xi_j$  is a zero-inflation parameter controlling the proportion of zero in the  
61 data,  $\sigma_j^2$  is the observation variance when the catch is positive.

62 The probability to obtain catch  $y_i$  is modelled in 2 components:

- 63 • the probability to obtain a zero catch ( $y_i = 0$ ). It is modelled as a Bernoulli variable  
64 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  
65 a 0 value with a Poisson distribution of intensity  $e^{\xi_j} \cdot \mu_j(x_i, t_i)$ . Then the probability  
66 to obtain a positive catch is given by  $1 - p_i$ .
- 67 • the value of the positive catch is modelled through a lognormal distribution L with  
68 expected value  $\frac{\mu_j(x_i, t_i)}{(1 - p_i)}$  and observation error  $\sigma_j^2$ . The standardization by  $(1 - p_i)$   
69 allows to keep the expectancy of the observation model to  $\mu_j(x)$ .

70 The catchabilities  $q_j$  are not identifiable per se and some additional constraints need to  
71 be set to estimate the relative catchability of each fleet (Alglave et al., (2022)). As is

72 common in variance analysis, one fleet catchability is set as reference level (e.g.  $q_{ref} =$   
73 1, here OTB\_DEF was used as the reference fleet) and the other fleets' catchabilities are  
74 estimated relatively to the reference fleet through the equation:

75 
$$q_j = k_j \cdot q_{ref}$$

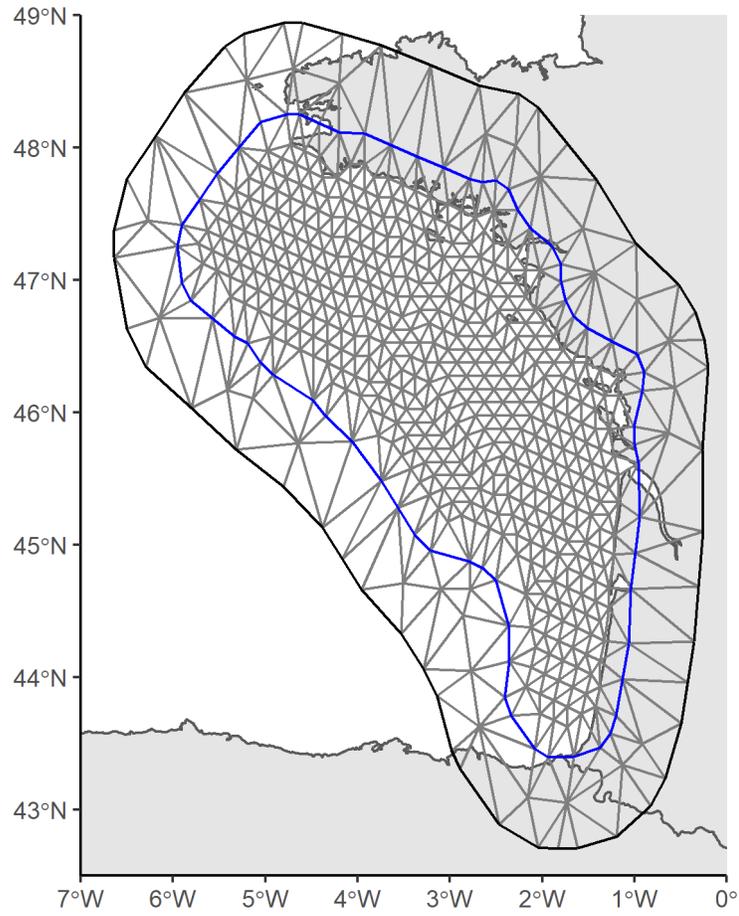
76

## 77 **5 THE SPDE APPROACH**

78 Estimating  $\Sigma$  is performed through the SPDE approach, which proves efficient to estimate  
79 correlation among points when the size of the covariance matrix becomes large.

80 Estimating  $\Sigma$  at every spatial locations can be computationally challenging when the  
81 dimension of  $\Sigma$  increases. A solution to overcome this computational burden was provided  
82 by Lindgren et al. (2011) through the SPDE approach. Instead of modelling the random  
83 effect as a GRF, the random effect  $\omega$  is represented as a Markovian representation of the  
84 GRF (GMRF) on the nodes of a triangulated mesh (e.g. see **Figure S5**).  $\omega \sim (0, Q^{-1})$ , with  
85  $Q$  the precision matrix which benefits from the sparse property of GMRF. The link between  
86 the random effect values estimated at each mesh node and the observations on the  
87 continuous space (defined on  $D \subset R^2$ ) is realized through linear interpolation. For  
88 extended details on the SPDE approach, GRF and the Matérn function, refer to Lindgren  
89 et al. (2011) and Cameletti et al. (2013).

90 To compute fine-scale spatial predictions of the biomass field, the latent field values  
91 obtained at the mesh nodes are interpolated on a discrete grid with much finer resolution  
92 (see **SM2**, **Figure S2**) as done in other packages such as VAST (Thorson, 2019).



93

94

95

96

**Figure S5.** Mesh used to estimate the spatio-temporal random effects  $\epsilon(x, t)$  and  $\eta(x, t)$ .

## 97 **6 ESTIMATING THE POINT PROCESS**

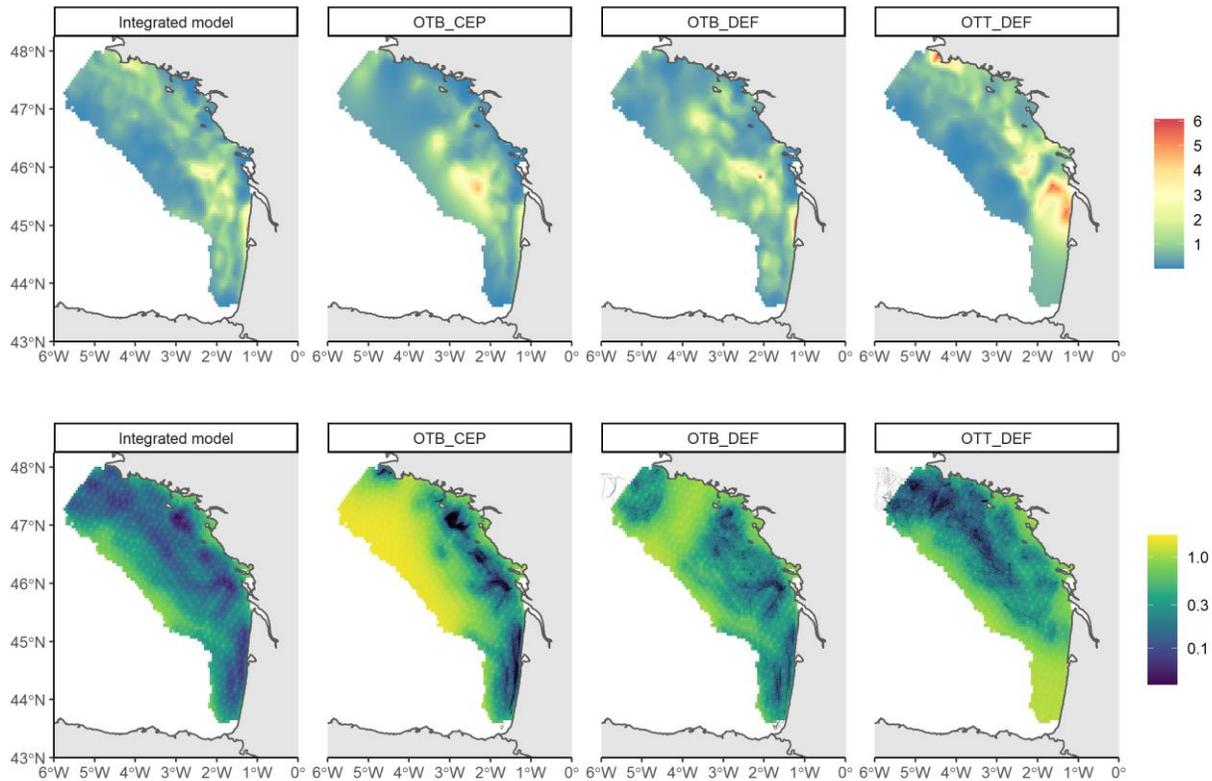
98 The log-likelihood of the point process is expressed as  $\log(\pi[X_{comj}]) = \sum_{i=1}^m \log \lambda_j(x_i) -$   
99  $\int_D \lambda_j(x) dx$  with  $m$  the number of observations. The term  $\int_D \lambda_j(x) dx$  (also called the  
100 'normalization constant') cannot be calculated explicitly and must be computed  
101 numerically (Renner et al., 2015). A common procedure consists in integrating  $\lambda_j(x, t)$   
102 over the study domain through a method referred as 'quadrature'; a set of 'quadrature  
103 points' are selected on the area and are used to approximate the log-likelihood through a  
104 weighted sum of the intensity  $\lambda_j$  at each quadrature points. Concretely, the log-likelihood  
105 is re-expressed as  $\log(\pi[X_{comj}]) \approx \sum_{i=1}^m \log \lambda(x_i) - \sum_{k=1}^n w_k \cdot \lambda(x_k)$  with  $w = \{w_1, \dots, w_n\}$   
106 the quadrature weights and  $x_k, k \in \llbracket 1, n \rrbracket$  the related quadrature points. For simplicity, we  
107 opted for a fine regular quadrature which makes all  $w_k$  equal (the quadrature points are  
108 the centroid of the cell grid of **Figure S2**). Note that more time-efficient methods exist  
109 (Jullum, 2020; Simpson et al., 2016), but will not be explored here as the one proposed  
110 by Renner et al. (2015) is a simple, stable and standard method for estimating the  
111 normalization constant.

112

## 113 **7 MAXIMUM LIKELIHOOD ESTIMATION**

114 As the random effects are in the logscale, the estimates of the biomass field and sampling  
115 intensity may be biased. We used the epsilon bias-correction to mitigate this bias (Thorson  
116 and Kristensen, 2016). Standard deviations were computed through the  $\delta$ -method.  
117 However, both methods imply an increasing computation time as number of time steps  
118 and the number/size of random effects increases. For this reason, bias correction and  $\delta$ -  
119 method were performed only for the biomass field values we explicitly map in the results.  
120 Still, fitting the model on the full time series can be computationally intensive. This is  
121 particularly true when the number of time steps and the number of fleets increases. To  
122 overcome this issue, we considered each year as a block and fitted each year (12 time  
123 steps) separately before merging the outputs of each block to reconstruct the full time  
124 series. We used the Datarmor supercomputer (Ifremer, 115 Gb available for each node,  
125 sequential fitting) to fit each block.

126 **8 BIOMASS PREDICTIONS AND RELATED COEFFICIENT OF VARIATION FOR**  
127 **NOVEMBER 2018**



128

129 **Figure S8.** Sole case study. Biomass field spatial predictions (top) and related  
130 coefficient of variation (bottom) obtained for November 2018. Integrated model: model  
131 combining all data sources. OTB\_CEP: model fitted to OTB\_CEP fleet. OTB\_DEF:  
132 model fitted to OTB\_DEF fleet. OTT\_DEF: model fitted to OTT\_DEF fleet. Black dots:  
133 fishing pings of each fleet.

134

135 Integrating the data from all the fleets allows a better coverage of the whole area and  
136 provide more accurate predictions on the full study domain. Predictions based on single-  
137 fleet models have high standard errors outside the fleet's sampling area while, when all  
138 fleets are integrated, standard deviation is drastically reduced in these areas (**Figure S8**).

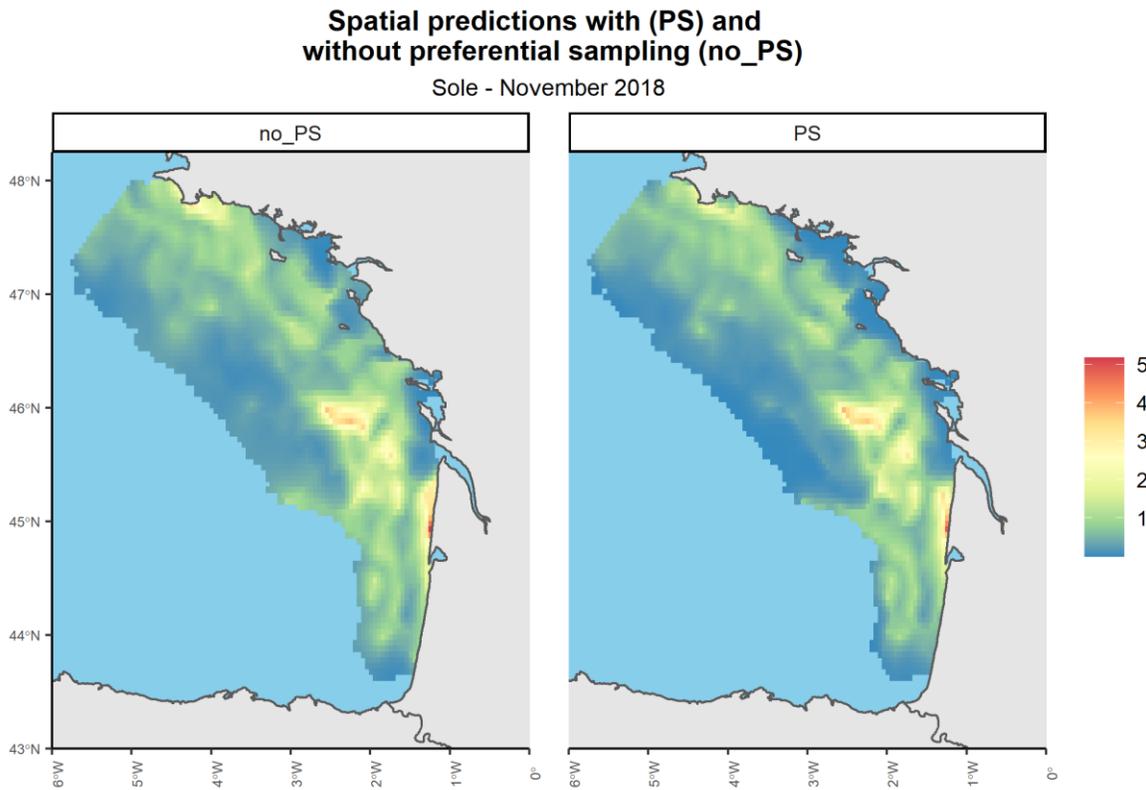
139

140

141

142 **9 SPATIAL PREDICTIONS WITH AND WITHOUT PS (NOVEMBER 2018)**

143

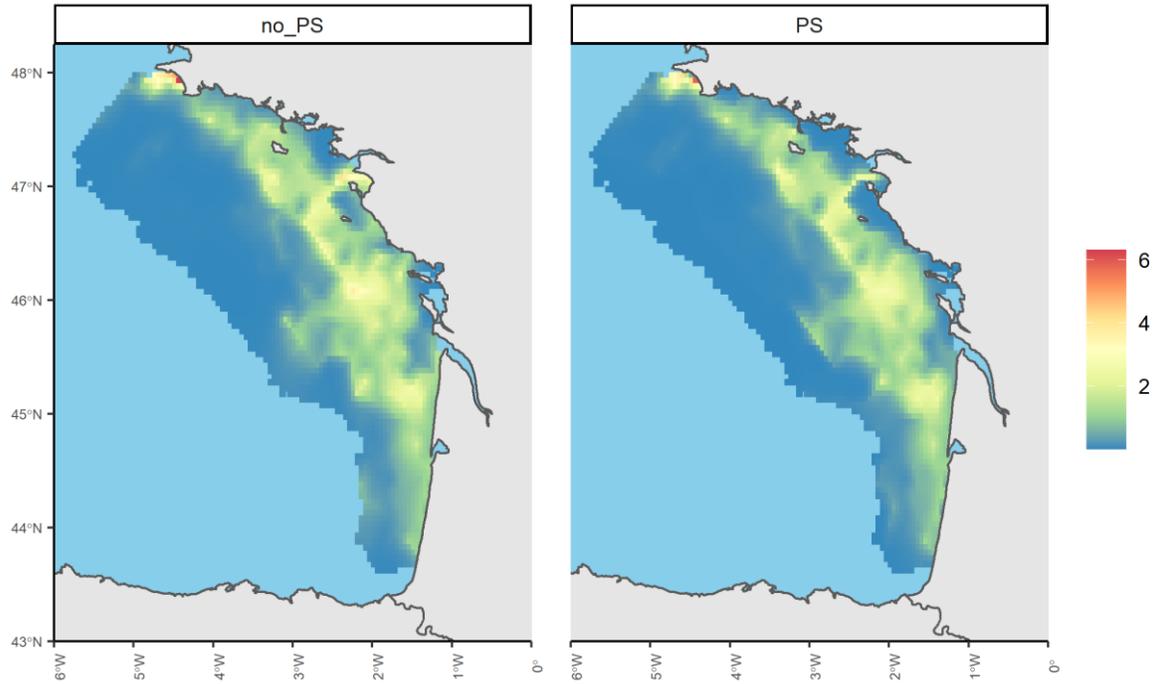


144

145 **Figure S9.1.** Sole. Biomass predictions in November 2018 for an integrated model  
146 accounting for PS or not (no\_PS).

**Spatial predictions with (PS) and  
without preferential sampling (no\_PS)**

Whiting - November 2018



147

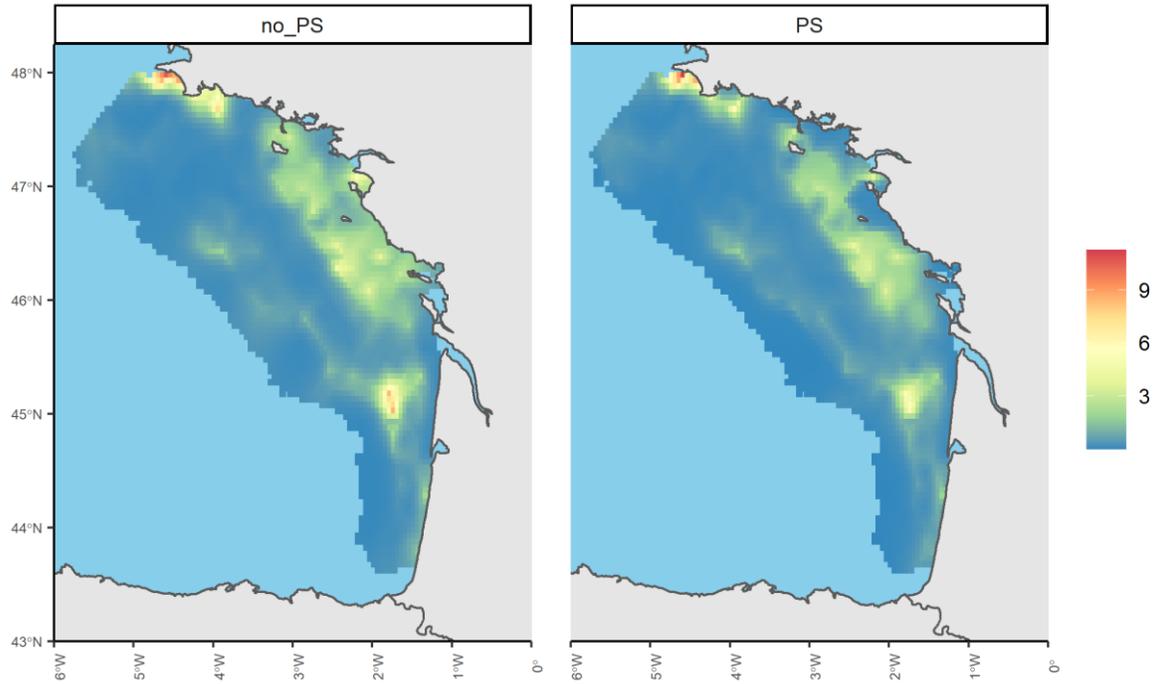
148

149

**Figure S9.2.** Whiting. Biomass predictions in November 2018 for an integrated model accounting for PS or not (*no\_PS*).

**Spatial predictions with (PS) and  
without preferential sampling (no\_PS)**

Squids - November 2018



150

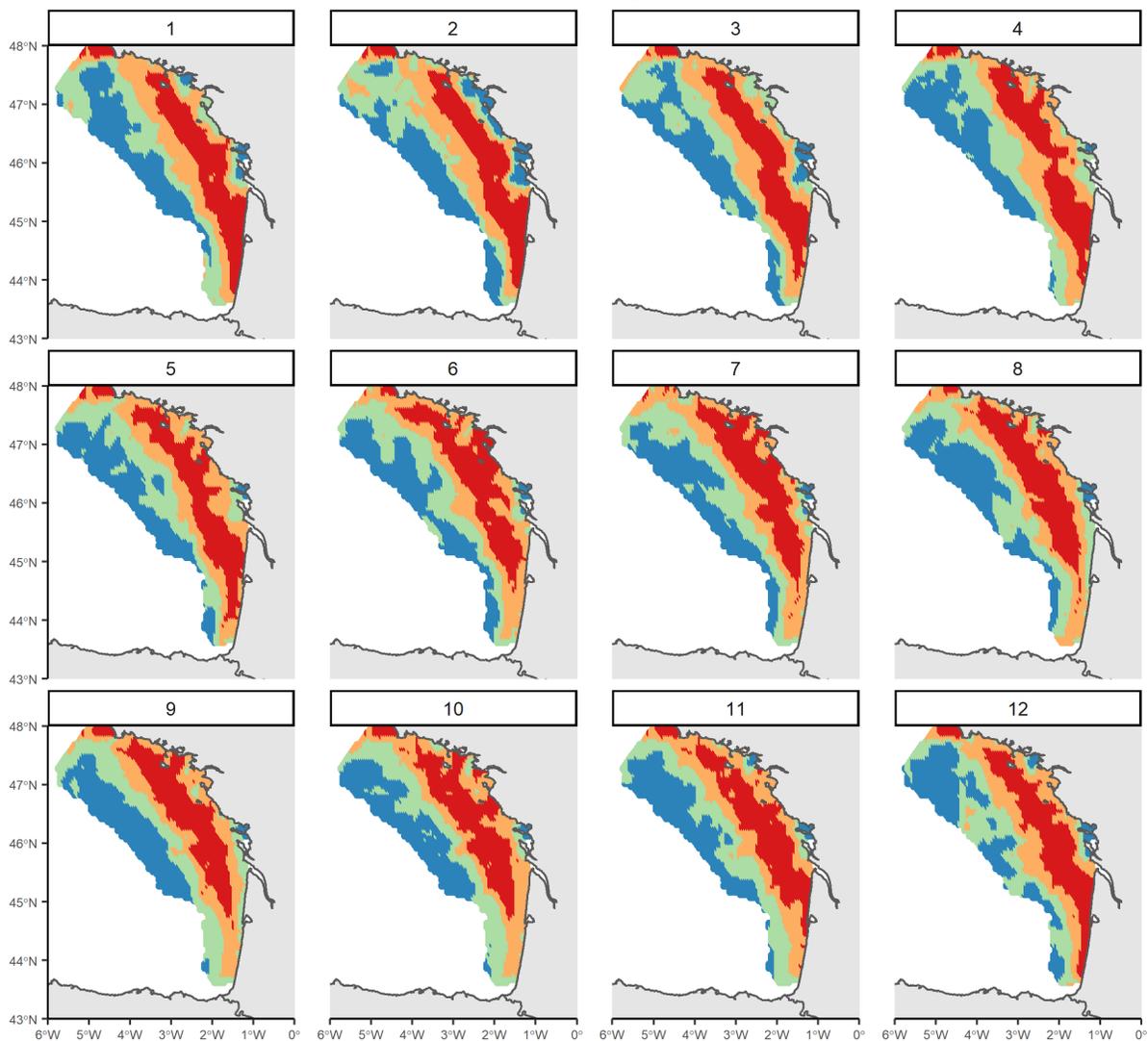
151 **Figure S9.3.** Squids. Biomass predictions in November 2018 for an integrated model  
152 accounting for PS or not (no\_PS).

153

154

155 **10 MONTHLY AVERAGE BIOMASS PREDICTIONS**

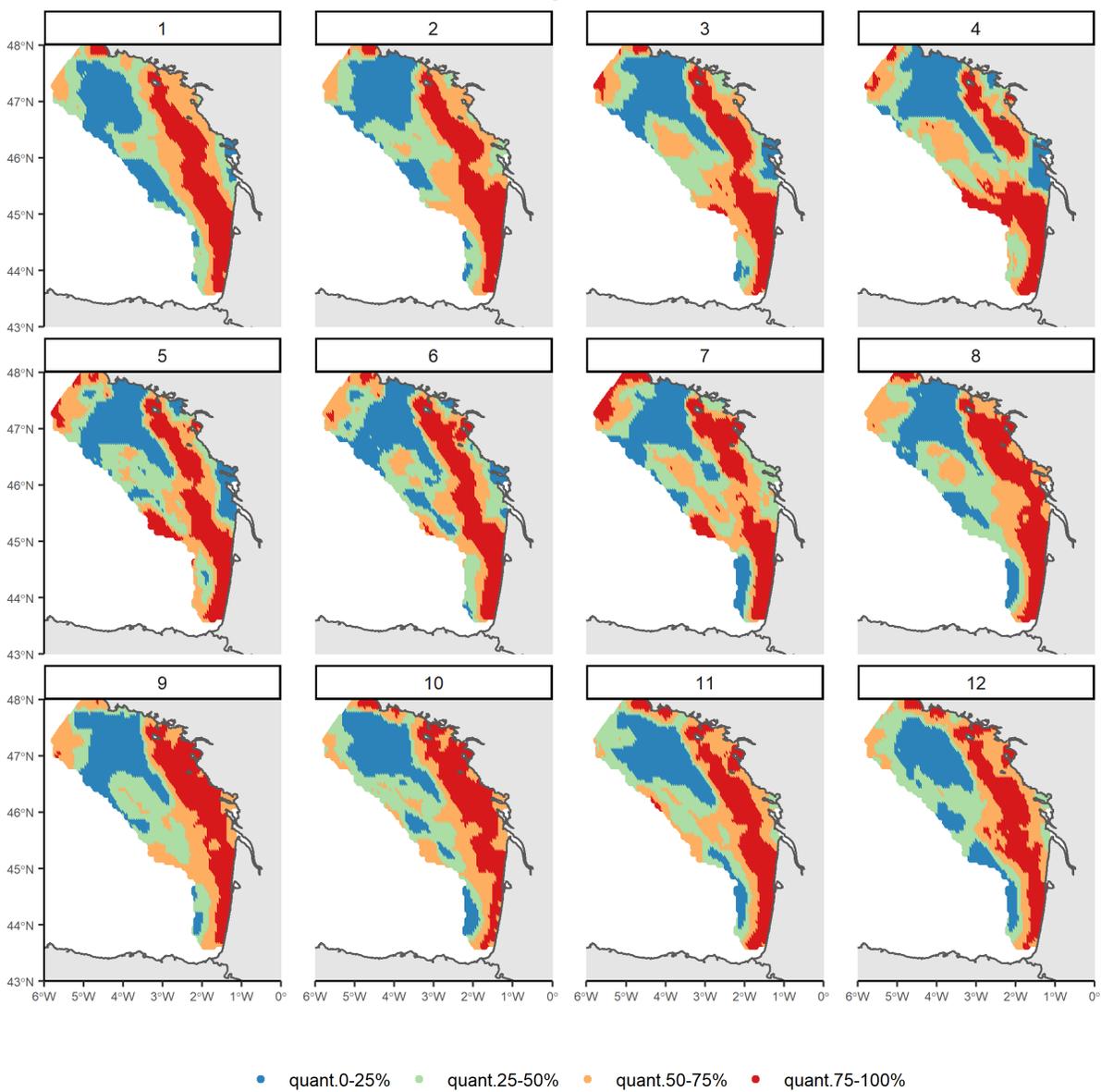
**Whiting**



● quant.0-25% ● quant.25-50% ● quant.50-75% ● quant.75-100%

156  
157 **Figure S10.1.** Whiting. Monthly biomass distribution averaged on the full period. Only  
158 quantile values are represented.  
159

### Squids



160

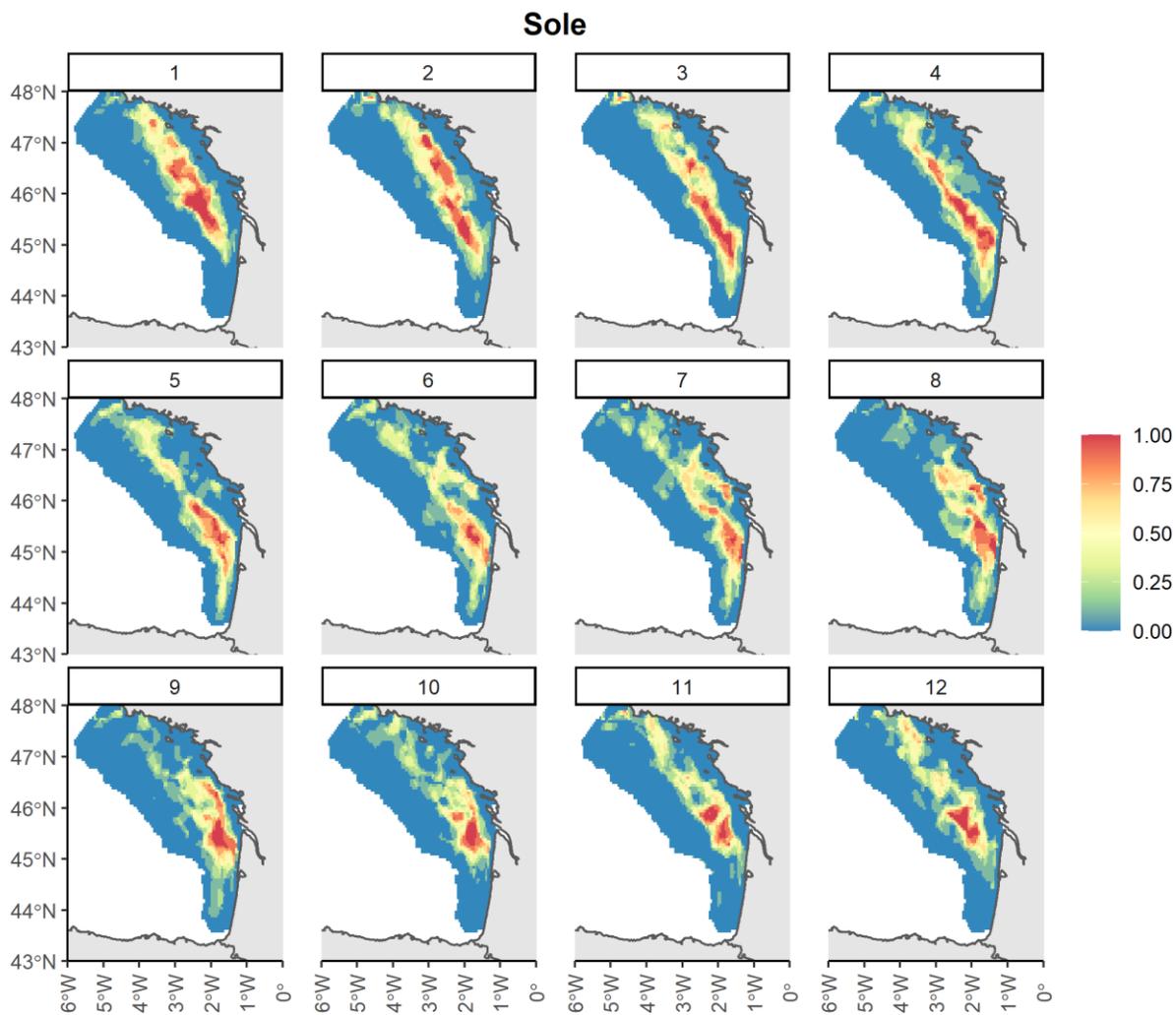
161

162

163

**Figure S10.2.** Squids. Monthly biomass distribution averaged on the full period. Only quantile values are represented.

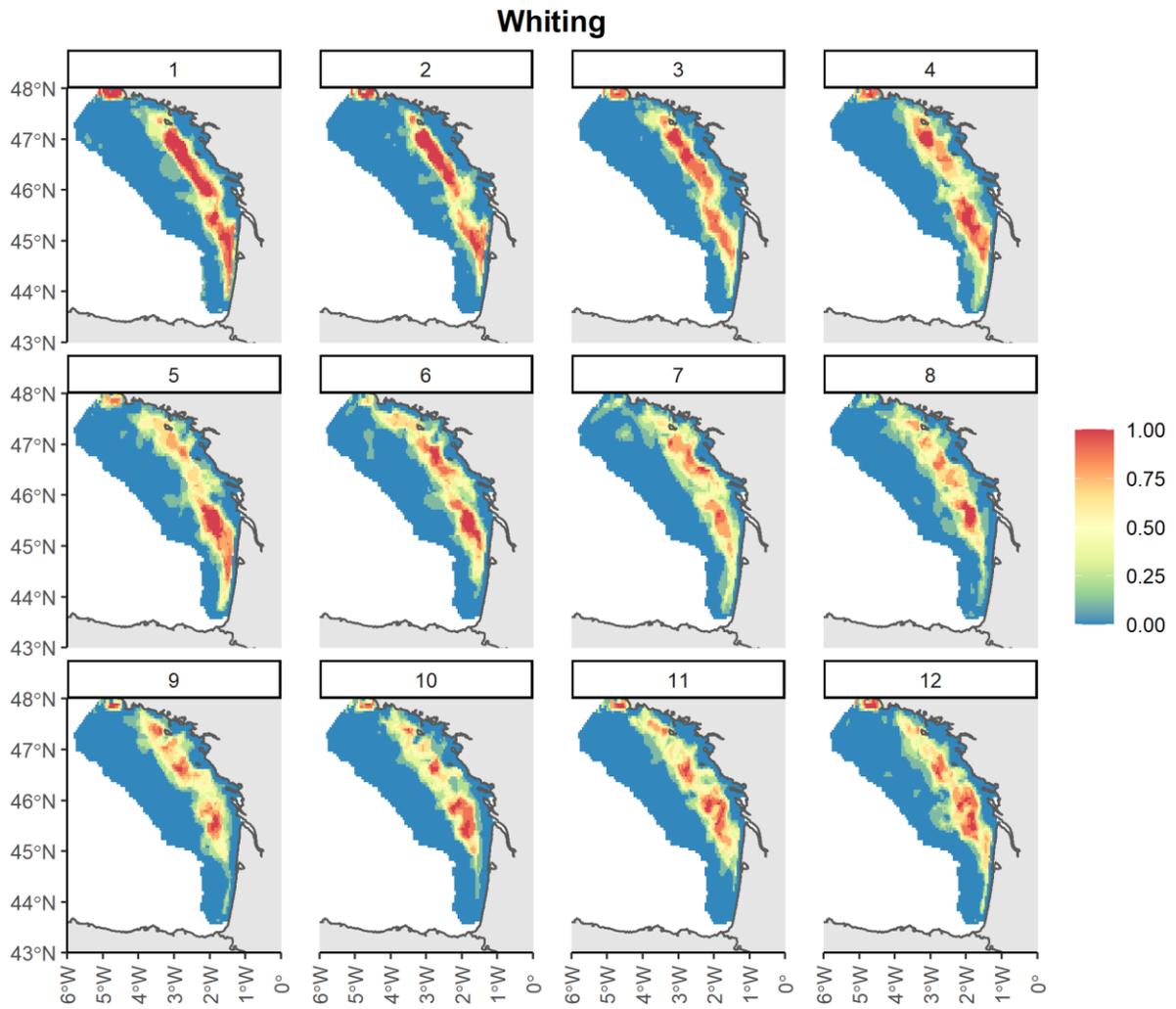
164 **11 PERSISTENCE INDEX MAPS**



165

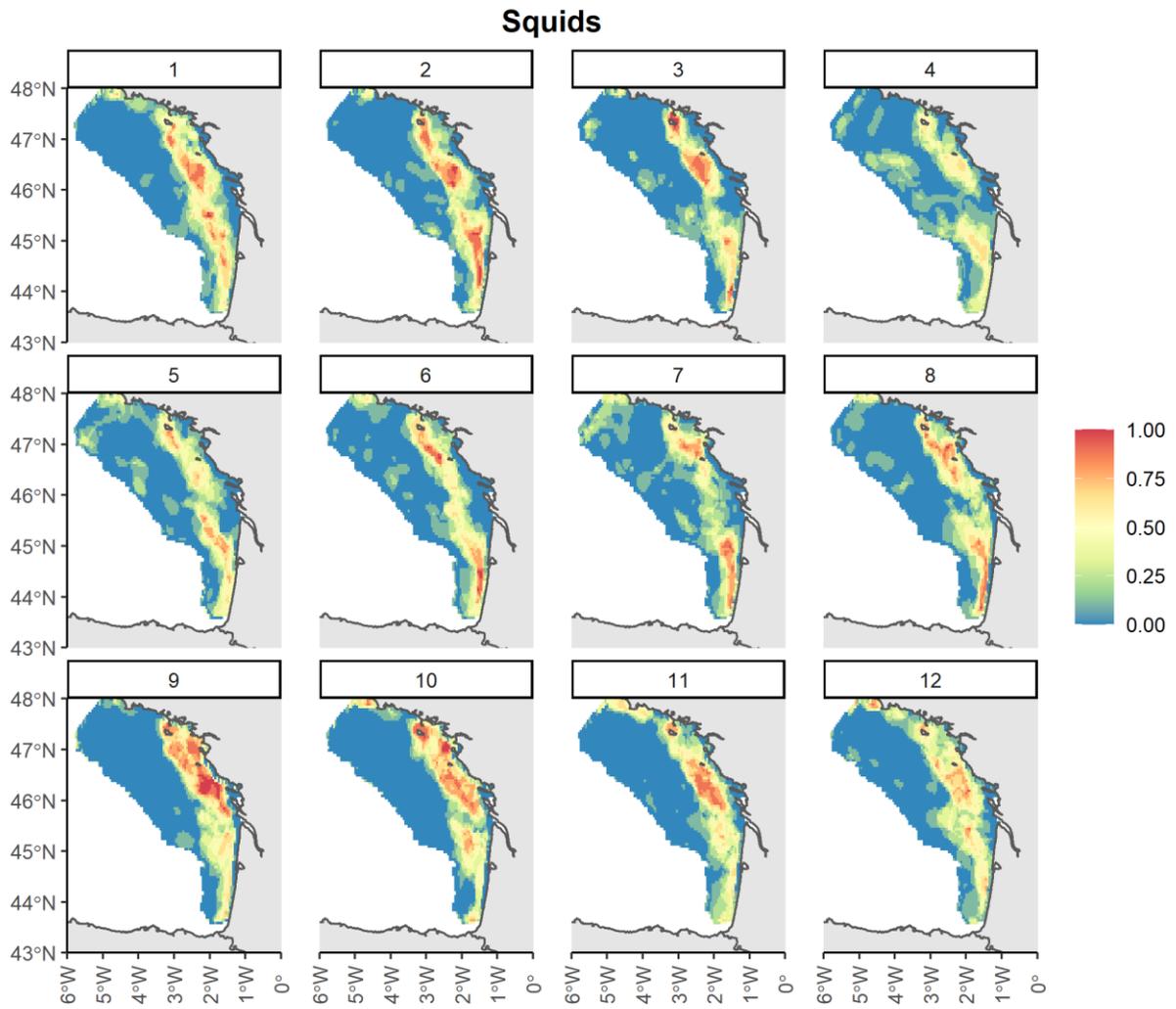
166 **Figure S11.1.** Sole. Monthly persistence indices. Aggregation over 2010-2018.

167



168

169 **Figure S11.2.** *Whiting. Monthly persistence indices. Aggregation over 2010-2018.*



170

171

**Figure S11.3.** Squids. Monthly persistence indices. Aggregation over 2010-2018.

172