

DISCUSSION

E.1 The value of ‘VMS x logbook’ data to explore fish spatio-seasonal patterns and species phenology at fine spatio-temporal scale

E.1.1 Material and methods

Case studies

We selected as case studies.:

- sole in the Bay of Biscay
- sea bass in the Bay of Biscay
- hake in both the Bay of Biscay and the Celtic Sea

These three case studies illustrate the progressive development of the use of catch declarations data to describe fish phenology and identify fish essential habitats.

First, sole represents a case where poor information are available to map fish distribution throughout the year. Information on spawning areas are available from an old egg and larvae survey (Arbault, Camus, and Bec, 1986) and a yearly beam trawl survey (Orhago) occurs each November and allows to sample few observations on the population (about 50 per year) to compute abundance index for the stock assessment (Biais, 2003). However, no additional data is available to map intra-annual species distribution and the knowledge of species distribution throughout the year remains very limited.

Second, hake illustrates the case where the analysis of logbooks has enabled to investigate the spatio-temporal distribution of mature individuals throughout the year at the scale of ICES rectangles. These were interpreted in terms of fish phenology and reproduction migrations (Poulard, 2001). Furthermore, studies investigating the spatial distribution of mature individuals during reproduction are available to identify potential reproduction areas (Tidd and Warnes, 2006).

More recently, Dambrine et al. (2021) characterized spawning areas of seabass at fine spatial scale based on logbooks data combined with VMS data. They identified aggregation areas from January to April based on occurrence data and interpreted these as reproduction areas. However, they only focused their analysis on the reproduction period and they did not investigate the whole year spatio-temporal variability.

Aim of the analysis

To go one step further, we propose to investigate the spatio-temporal variability of these species throughout the year based on monthly species distribution maps built on ‘VMS x logbooks’ data. Our goal is to provide a generic and synthetic approach to analyze the key spatio-temporal patterns that structure species distribution on the full year. We aim at identifying potential long-term trends in species distribution, seasonal phases that structure species distribution and eventually the punctual events that affected species distribution in the previous years. Our final goal is to interpret these signals in terms of fish phenology and then to identify some key functional zones of the species (i.e. fish essential habitats).

Data and model

The spatio-temporal models were fitted between 2008 and 2018. We filter the mature fraction of the landings to map the potential breeders of the population and interpret the model outputs in terms of reproduction phenology.

The models integrate several trawl fleets that ensure good coverage of the area. For hake we selected 3 bottom trawl fleets targeting demersal fish (OTB_DEF_>=70_0, OTB_DEF_100_119_0, OTB_DEF_70_99_0) and one otter trawl fleet targeting demersal fish (OTT_DEF_>=70_0); for sole we selected a bottom trawl targeting demersal fish (OTB_DEF_>=70_0), a bottom trawl targeting cephalopods (OTB_CEP_>=70_0), and an otter trawl targeting demersal fish (OTT_DEF_>=70_0); for seabass, one bottom trawl targeting cephalopods (OTB_CEP_>=70_0) and demersal fish (OTB_DEF_>=70_0) and one pelagic trawl fleet targeting demersal fish (PTM_DEF_>=70_0).

Preferential sampling was not accounted for in inference as this does not modify the overall patterns of species distribution and it would lead to increased computation time for little gain regarding spatial predictions (Cf. chapter 4).

Decomposing model outputs: Empirical Orthogonal Factors (EOF)

Base formulation of EOF

Empirical Orthogonal Functions is a method that has been developed by Lorenz (1956) for weather forecasting applications. The original aim of the technic was to deduce from a set of spatio-temporal maps a smaller set of maps that best describe and summarize the spatio-temporal process of interest. The main idea is to define the spatio-temporal process $S(x, t)$ as a linear combination of spatial patterns \mathbf{p}_m (named EOF) each related to temporal indices $\alpha_m(t)$ (Equation E.1). These temporal indices indicate when the spatio-temporal process is distributed following their related spatial pattern.

$$S(x, t) = \sum_{m=1}^M \alpha_m(t) \cdot p_m(x) + r^M(x, t) \quad (\text{E.1})$$

where $r^M(x, t)$ is the residual variation not captured by the M modes of variation $\alpha_m(t) \cdot p(m, x)$. $x \in \llbracket 1, n \rrbracket, t \in \llbracket 1, T \rrbracket$.

To estimate the temporal components $\alpha_m(t)$ and the spatial patterns $p(m, x)$, some criteria need to be set. A natural choice is to minimize the residual variation to best capture the variability of $S(x, t)$ (i.e. $R^M = \sum_{m=1}^M (r_m^M)^2$ is minimized) and set orthogonal constraints between the modes of variability so that each mode is the 'best' representation of variability from a statistical point of view (Equations E.2, E.3). This last criteria is further discussed in the discussion.

$$\sum_{x=1}^N p_m(x) \cdot p_j(x) = \delta_{mj} \equiv \begin{cases} 1 & \text{if } m = j \\ 0 & \text{if } m \neq j \end{cases} \quad (\text{E.2})$$

$$T \cdot \overline{\alpha_m^* \alpha_j^*} = a_m \delta_{mj} \quad (\text{E.3})$$

with $a_m \geq a_{m+1} \geq 0$, $\overline{(\)}$ denoting the time average and $()^*$ a departure from the time average.

Matrix formulation

Let's write these equations in matrix terms by introducing the $T \times N$ matrix \mathbf{S} , \mathbf{S}^* ,

\mathbf{Q} , \mathbf{Q}^* . They respectively refer to $S(x, t)$, $S^*(x, t)$, $\alpha_m(t)$ and $\alpha_m^*(t)$. We denote \mathbf{Y} as a square matrix of order N with elements $p_m(x)$. Then, the problem can be reformulated as:

$$\mathbf{S} = \mathbf{Q}\mathbf{Y} \quad (\text{E.4})$$

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{I} \quad (\text{E.5})$$

$$\mathbf{Q}^{*T}\mathbf{Q} = \mathbf{D} \quad (\text{E.6})$$

with $(\cdot)^T$ the transpose, \mathbf{I} the identity, \mathbf{D} a diagonal matrix with diagonal being equal to a_m/T .

Finally, by introducing $\mathbf{A} \equiv \mathbf{Q}^{*T}\mathbf{Q}^*$ (whose elements are proportional to the covariance of $\alpha_m(t)$), we can rewrite these equations as:

$$\mathbf{Y}\mathbf{A}\mathbf{Y}^T = \mathbf{D} \quad (\text{E.7})$$

This is a standard "eigenvalue-eigenvector" problem. As a consequence, a link can be done with standard multivariate analysis such as PCA.

Thus, in addition to the spatial patterns (and the related temporal components) that appear in the EOF formulations (Equation E.1) and that can be obtained by diagonalizing the problem in Equation E.7, additional analysis can be performed to obtain similar visualization as in PCA (e.g. plot of individuals and variables, contribution of locations and time steps to the several dimensions) possibly coupled with standard clustering analysis.

Analysis

We filter the number of dimensions for each species based on the graph of the variance captured by each dimension (Figure E.1). As is commonly done in PCA analysis, we cut the graph at the dimension where there is a drop in the variance explained. For sole, we filtered the six first dimensions. For seabass, we select the first dimension only. For hake in the Bay of Biscay as well as for hake in the Celtic sea, we filtered the two first dimensions.

Then, we analyze the spatio-temporal model outputs by presenting the spatial patterns $p_m(x)$ (which are the eigen-vectors in the diagonalization). These can be seen either as

maps that structure \mathbf{S} (e.g. Figure E.2) or either as standard variable plot as in classical PCA analysis (Cf. Figure E.6, center figure).

The temporal index $\alpha_m(t)$ can be either seen as time series (e.g. Figure E.2) or as a classical plot of individuals on the first components of the PCA (Cf. Figure E.6, left figure).

When confronted together, the spatial representations and the related temporal indices inform how the spatial patterns that structure species distribution vary in time. When the loading factor $\alpha_m(t)$ are positive (resp. negative), then species distribution \mathbf{S} at time step t is distributed following the EOF map $p_m(x)$ (resp. $-p_m(x)$). Then the evolution of $\alpha_m(t)$ tells how \mathbf{S} evolve in time. Such patterns can be interpreted in regards to the spatial ecology of the species.

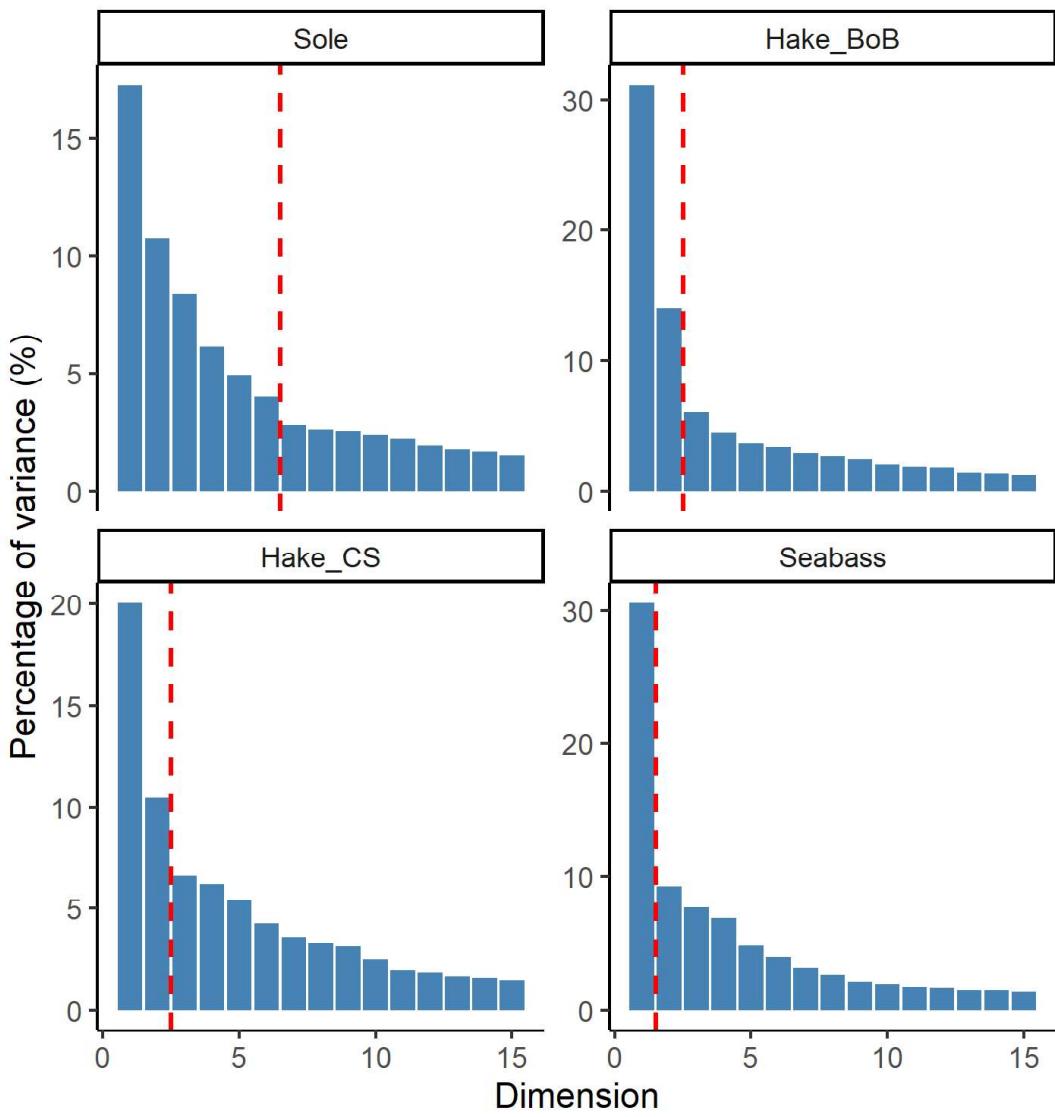


Figure E.1 – Graph of variance explained by each dimension for each species. Dashed line: cutting level of the EOF.

E.1.2 Results

Spatio-temporal representation

All species evidence clear seasonal patterns (Figure E.2, E.3, E.4, E.5); In all cases, the 2 first EOF dimensions capture a seasonal signal.

For sole, some periodic signal is evidenced highlighting high biomass in offshore areas in winter (December to April) and relatively coastal distribution in summer. This is consistent with the results highlighted in chapter 4 (see section D) and in Arbault, Camus, and Bec (1986) where sole is evidenced to reproduce in relatively offshore areas.

For seabass, a strong positive anomaly occurs seasonally each January/February. High biomass zones are mainly localized along and off shore the Vendée coast and up to the plateau of rochebonne (i.e. all the locations that are not blurred in Figure E.3). This is consistent with the paper from Dambrine et al. (2021) that emphasizes that both areas and months that we emphasize corresponds to the reproduction period and area of seabass. In addition to their work, we provide information on biomass densities (not only on occurrence density) which can help to order the aggregation areas of fish and to identify where most fish aggregate.

For hake, similar (but weaker) patterns can be evidenced on the first two EOF (Figure E.4). In the Bay of Biscay, hake has a more coastal distribution in winter (January/February/March) compared with summer. For hake in the Celtic Sea, there is a strong signal on the first EOF (Figure E.5) in the center of the Celtic Sea occurring in summer, a period where no information is available from the literature to inform which process could occur here. When looking at the second EOF some high biomass patterns come back seasonally on the Eastern of the Celtic Sea and match with the distribution of mature individuals from old surveys focusing on reproduction (Tidd and Warnes, 2006).

Astonishingly, the high biomass patterns on the Eastern of the Celtic Sea occur slightly later than in the Eastern of the Bay of Biscay. This could be related to a later reproduction period in the Northward areas due to temperature differences between these two areas.

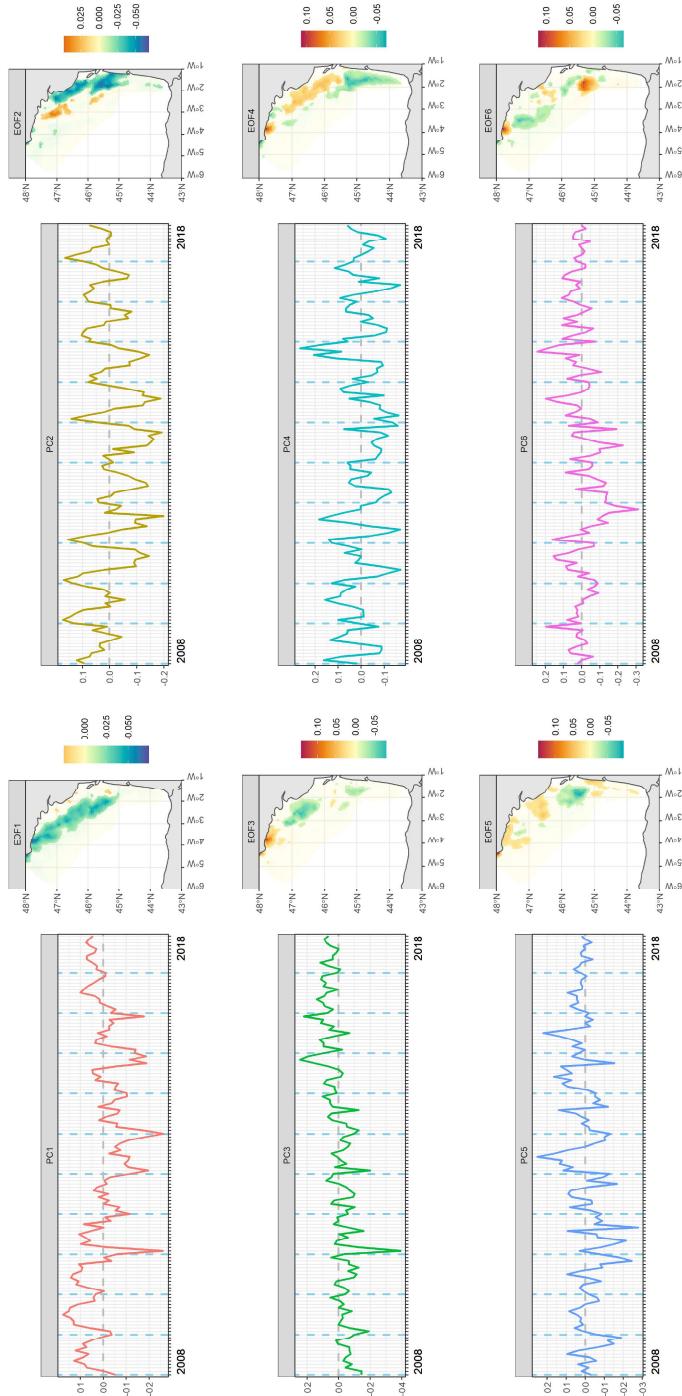


Figure E.2 – Sole. Six first EOF maps and time-series for sole. Blue dashed line: January. For EOF maps, the locations that do not have a significant contribution to the dimension are blurred.

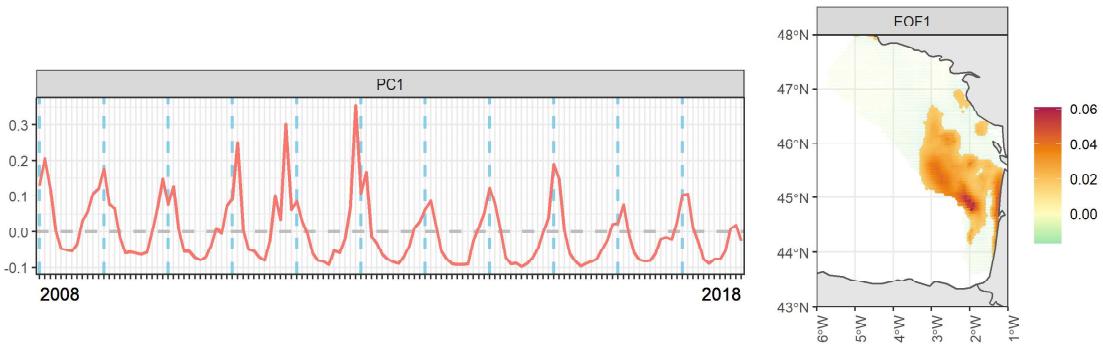


Figure E.3 – Sea bass in the Bay of Biscay. First EOF map and time-series. Blue dashed line: January. For EOF maps, the locations that do not have a significant contribution to the dimension are blurred.

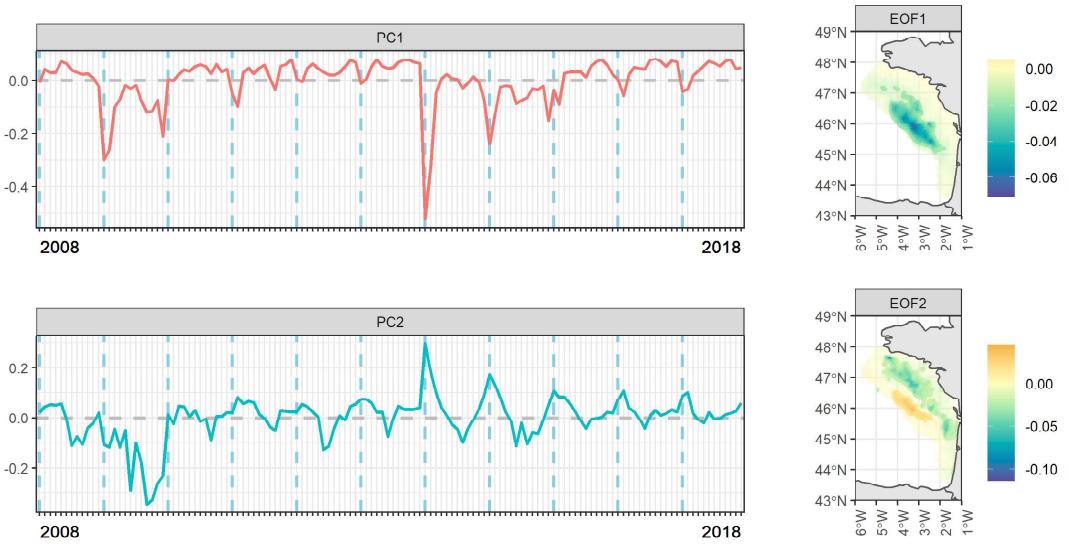


Figure E.4 – Hake in the Bay of Biscay. Two first EOF maps and time-series. Blue dashed line: January. For EOF maps, the locations that do not have a significant contribution to the dimension are blurred.

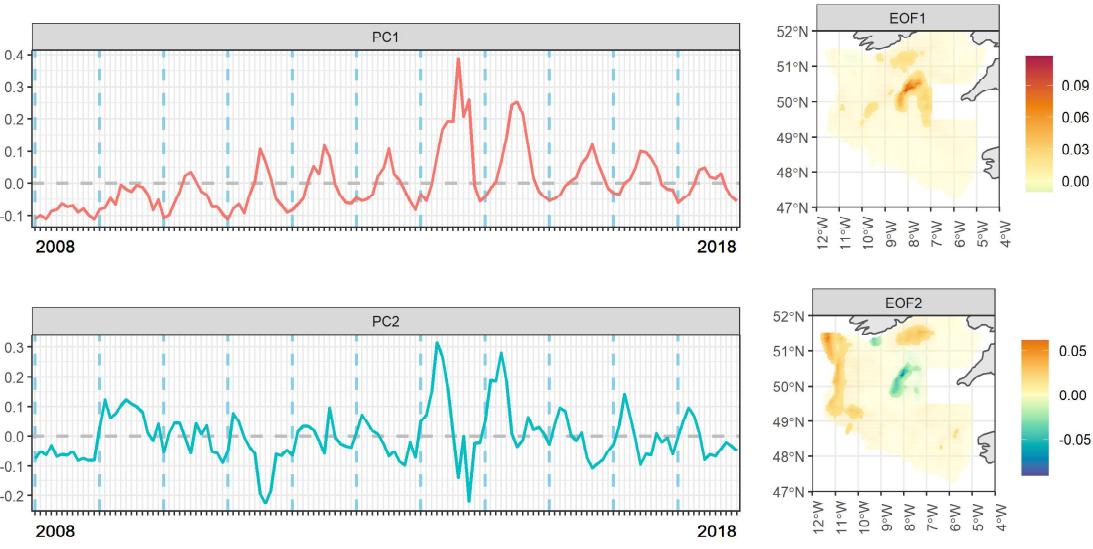


Figure E.5 – Hake in the Celtic Sea. Two first EOF maps and time-series. Blue dashed line: January. For EOF maps, the locations that do not have a significant contribution to the dimension are blurred.

PCA representation

The PCA representation of the spatial and temporal patterns coupled with hierarchical clustering allows to identify sets of points and time steps that can be interpreted as combination of seasons and functional zones. This is illustrated with the sole case study where Figure E.6 and E.8 highlights that 3 clusters of time steps can be grouped into 3 seasons:

1. the red cluster (Figure E.6, left) that regroups winter months (December to February - Figure E.7). This can be related to the cluster 1, 2 and 3 on the graph of locations (center figure) and maps of clusters (right figure) which corresponds to areas where sole reproduce (see chapter 4 and Arbault, Camus, and Bec (1986)).
2. the green cluster (Figure E.6, left) mainly corresponds to spring and early summer months (March to June - Figure E.7). Sole progressively go back to coastal areas.
3. the blue cluster (Figure E.6, left) mainly corresponds to autumn and spring distribution (July to November - Figure E.7). Sole has a more coastal distribution (biomass of sole are high in the blue area cluster on the right plot of Figure E.6)

and biomass also aggregates in the hotspots of the South of the Bay of Biscay (green cluster in the right plot of Figure E.6).

Then related 3 areas could be interpreted as functional zones (right plot in Figure E.6): (1) a reproduction area (the orange and yellow clusters) where biomass aggregates mainly from December to January, (2) a coastal area where biomass is high in autumn and winter (the blue cluster) and (3) an area where biomass is always high but specifically in autumn and winter (green cluster). These two last areas could be interpreted as feeding areas: sole feed in these areas during autumn and early winter before going back to reproduction grounds each year from December to February. The green cluster locations could also be interpreted as the arrival of juveniles in the mature population.

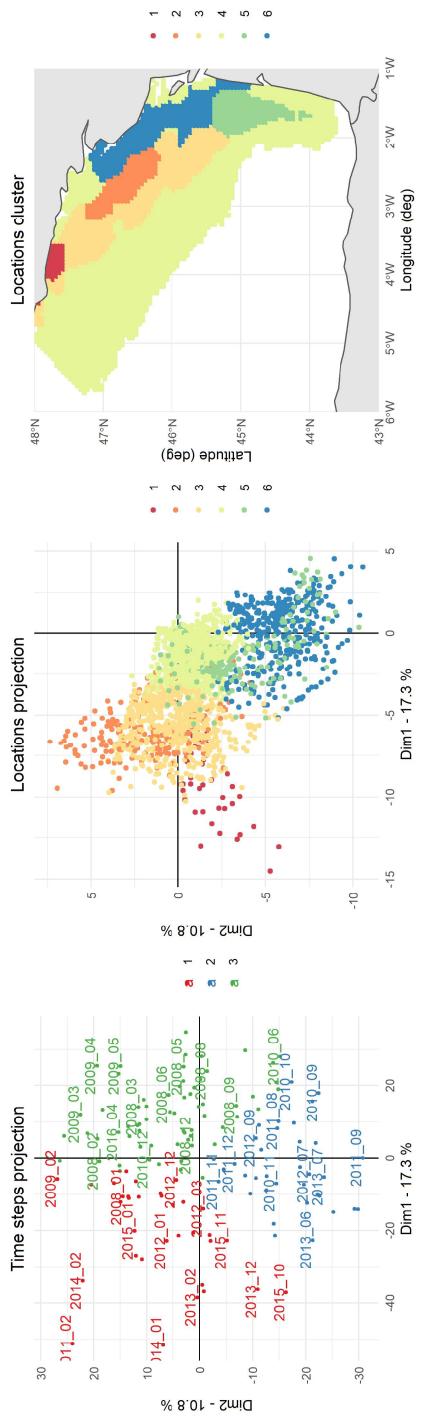


Figure E.6 – (Left) Projection of the time steps indices on the 2 first dimensions of the EOF. Color: cluster identified through HAC analysis. All the points are related to their time step. (Center) Projection of the locations on the 2 first dimensions. Color: cluster identified through HAC analysis. (Right) Spatial representation of the clusters.

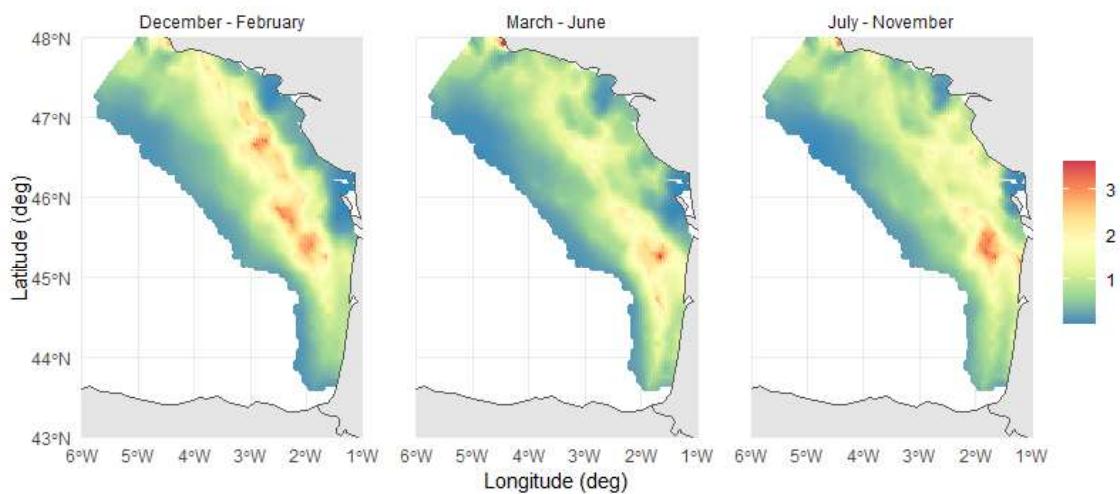


Figure E.7 – Mean pattern for each seasonal cluster identified in Figure E.6. December to February corresponds to the red cluster; March to June corresponds to the green cluster; July to November corresponds to the blue cluster.

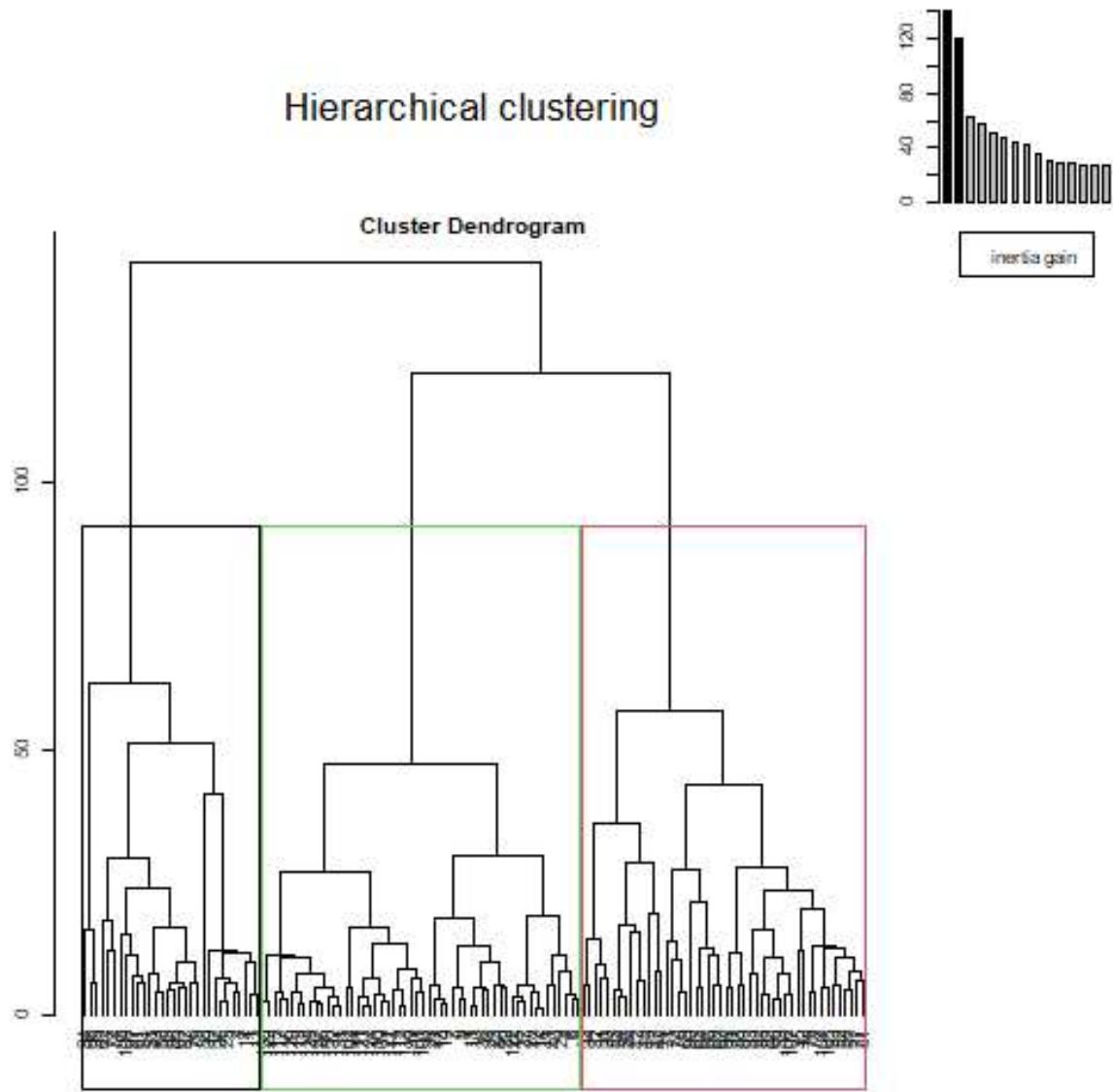


Figure E.8 – Sole. Clustering tree for the time steps.

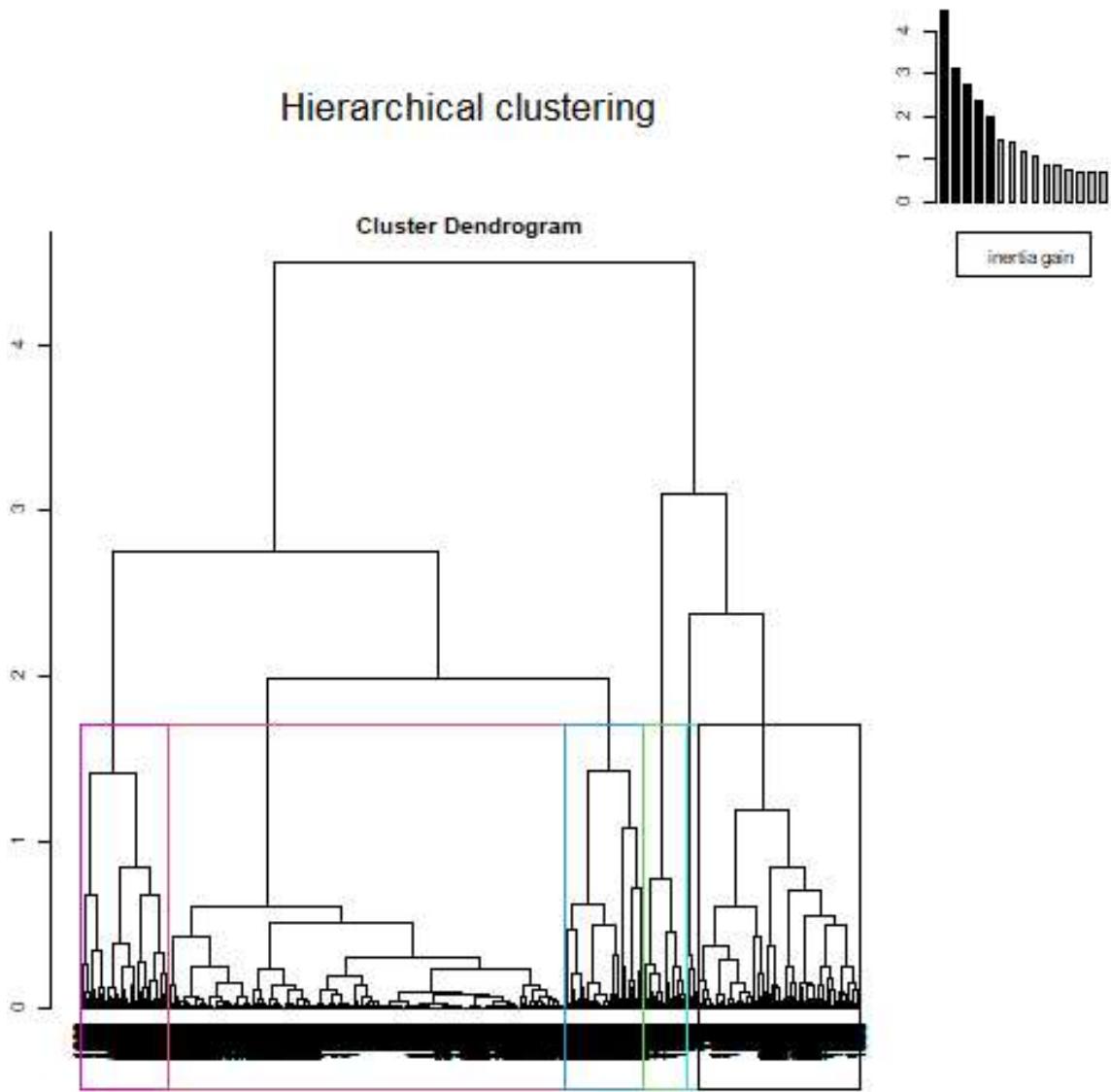


Figure E.9 – Sole. Clustering tree for the locations.

E.1.3 Discussion

- ‘VMS x logbooks’ provide information to infer species distribution at a monthly time step. This makes possible to identify spatio-seasonal patterns and to interpret these in terms of fish functional zones for sole. Similar results are available for the other species.

-
- Applying a similar approach for several species would enable to identify spatio-seasonal patterns that are common to these species and possibly identify essential habitats for a set of species.
 - Our approach raises methodological challenges as the signals of the different maps and time series are sometimes redundant (i.e. several time series emphasize the same seasonal pattern), but one would like to better synthesize the information of this signal in a single time-series and a single map. Methods exist to better disentangle the spatial signal from a set of maps. These are based on geostatistical methods e.g. MAF and EOM (Bez, Renard, and Ahmed-Babou, 2022). However, these are strongly sensitive to the choice of the neighborhood and they do not account for temporal correlation. Developing methods that better disentangle the signal without the limitations of MAF would be a valuable contribution to spatial statistics. This could be possible by modifying the constraints of EOF by specifying orthogonality constraints that accounts for spatio-temporal correlations.

