

# Investigating fish reproduction phenology and essential habitats by identifying the main spatio-temporal patterns of fish distribution

Baptiste Alglave <sup>1,2,3,\*</sup>, Maxime Olmos<sup>2</sup>, Juliette Casemajor<sup>2</sup>, Marie-Pierre Etienne<sup>4</sup>, Etienne Rivot<sup>3</sup>, Mathieu Woillez <sup>2</sup>, Youen Vermard<sup>5</sup>

<sup>1</sup>Lab-STICC (Information and communication technology research), Université Bretagne Sud, Vannes 56000, France

<sup>2</sup>DECOD (Ecosystem Dynamics and Sustainability), L'Institut Agro, INRAE, IFREMER, Plouzané 29280, France

<sup>3</sup>DECOD (Ecosystem Dynamics and Sustainability), L'Institut Agro, INRAE, IFREMER, Rennes 35042, France

<sup>4</sup>Mathematical Research Institute of Rennes IRMAR, Rennes University, Rennes 35042, France

<sup>5</sup>DECOD (Ecosystem Dynamics and Sustainability), L'Institut Agro, INRAE, IFREMER, Nantes 44980, France

\*Corresponding author. Lab-STICC (Information and communication technology research), Université Bretagne Sud, Vannes 56000, France. E-mail: [baptiste.alglave@univ-ubs.fr](mailto:baptiste.alglave@univ-ubs.fr)

## Abstract

Fish spawning phenology is a major concern for conservation and fisheries management. New intensive data sources, such as GPS-based tracking data and high-resolution catch declaration data, are becoming increasingly available in the field of marine ecology. These data benefit from high spatiotemporal resolution and open new research avenues for investigating the interannual variability in fish phenology. In this paper, we demonstrate how an integrated species distribution model informed by commercial catch data combined with spatiotemporal dimension reduction methods known as empirical orthogonal functions (EOFs) can be used to synthesize spatiotemporal signals in fish reproduction phenology. Specifically, we address the following questions: (1) Can we identify seasonal spatial patterns that can be interpreted in terms of reproductive phenology and essential habitats? (2) Can we identify changes in reproductive phenology over time? (3) Are these changes related to environmental drivers? The analysis illustrates the reproductive phenology of three key commercial species in the Bay of Biscay (sole, hake, and sea bass). The EOF analysis emphasized strong seasonal spatiotemporal patterns that correspond to reproduction patterns and feeding patterns. Based on this methodology, we identified seasonal variations in the timing of reproduction, and we related these variations to sea surface temperature, a key driver of fish reproduction.

**Keywords:** species distribution; spatiotemporal modeling; spawning phenology; spawning season; intra- and inter-variability

## Introduction

To complete their life cycle, fish require different habitats specific to different life stages (Harden 1969). These habitats, also known as essential fish habitats are associated with key demographic processes in the fish life cycle, such as spawning, feeding, and migration, and are characterized by a strong concentration of individuals within a spatially restricted area (Delage and Le Pape 2016). Fish may have to adapt to rapid environmental changes by physically adapting or by modifying their spatiotemporal behavior (spatial displacement or change in the seasonal timing of their demographic processes) to stay within the ideal habitat conditions.

Understanding changes in the phenology of demographic processes is critical for the management of fish populations. Knowledge of seasonal habitat use, the timing of migrations and the location of spawning areas is key for preserving fish habitats and ensuring the renewal of marine resources (Lieth 2013, Delage and Le Pape 2016). Reproduction is a critical stage of the life cycle, and knowledge of the areas where fish spawns and the timing of spawning require specific attention in terms of spatial planning and fisheries management (Grüss et al. 2019, Biggs et al. 2021).

Most of the available information available for investigating fish spatiotemporal demographic processes such as reproduction comes from data with only sparse spatiotemporal coverage (e.g. scientific survey data, capture-mark-recapture tagging data). Typically, such tagging methods involve capturing individuals from a population, marking them in a non-invasive manner, releasing them back into the environment, and then recapturing a sample of individuals (Jolly 1965, Seber 1965). These methods can be very efficient at inferring the movement of fish (de Pontual et al. 2023, Randon et al. 2021). However, the methods can be costly, and few data can be recorded. Typically, scientific surveys usually occur once a year and provide samples only on the time span of the survey, which does not necessarily depend on the timing of reproduction (Alglave et al. 2023). Onboard observer data provide additional data for the whole year by recording fishing catches on a small portion of commercial fleets (Rufener et al. 2021). With these data, it is possible to infer fish distributions seasonally or quarterly (Kai et al. 2017, Olmos et al. 2023). However, this temporal resolution is generally not sufficient to precisely investigate the phenology of demographic processes such as reproduction that occur at

shorter temporal scales e.g. months and weeks (Biggs et al. 2021).

Methods to combine fishers' declarations (logbooks) with vessel monitoring system (VMS) data ("VMS  $\times$  logbooks" hereafter) have recently been developed to provide fine-scale information on fishing activity and fishing landings (Bastardie et al. 2010, Hintzen et al. 2012). Specifically, "VMS  $\times$  logbooks" data sources have been used to infer the spatiotemporal distribution of fish at a fine scale (Murray et al. 2013, Azevedo and Silva 2020, Dambrine et al. 2021, Alglave et al. 2022). These data benefit from a high spatiotemporal resolution and open large research avenues for investigating the long-term and short-term variability of fish spatial distributions.

Recently, a modeling framework has been built to infer fish spatial distributions by integrating "VMS  $\times$  logbooks" data from distinct fishing fleets while handling preferential sampling of fisheries data (Alglave et al. 2022). The framework has been extended in time using a monthly time step. It has been applied to map fish aggregation areas to identify spawning grounds for a few key species of the Bay of Biscay (Alglave et al. 2023). Nevertheless, these approaches face the difficulty of simultaneously interpreting both interannual variability and intra-annual (seasonal) variability in fish spatiotemporal distribution and hence only investigate a single year of data (Azevedo and Silva 2020) or a specific period over several years (Alglave et al. 2023). Consequently, these studies do not take advantage of the large amount of information comprising the inter- and intra-annual variability of a long-term time series.

Spatiotemporal dimension reduction techniques such as empirical orthogonal functions (EOF; Lorenz 1956, Hannachi et al. 2007) can provide insights into the spatiotemporal variability of fish population processes. EOF is derived from principal component analysis (PCA) and, in most cases, it has been used to characterize physical oceanography conditions (e.g. Woillez et al. 2010). Some recent studies have investigated fish processes using EOF (Petitgas et al. 2014, Thorson et al. 2020a, 2020b, Grüss et al. 2021). However, these studies only aimed to synthesize the interannual variability in ecological processes and ignored the seasonal (intra-annual) variability.

In this paper, we aimed to demonstrate the potential of integrated spatiotemporal hierarchical models (ISTHMs; Alglave et al. 2022, 2023) combined with EOF to achieve the following:

- (i) Identify seasonal spatiotemporal patterns in fish distributions that can be interpreted in terms of essential habitats. A specific focus is given to reproduction phenology.
- (ii) Infer interannual changes in reproductive phenology over long time series.
- (iii) Assess the link between environmental drivers and the control of phenology.

The framework is illustrated with sole (*Solea solea*), hake (*Merluccius merluccius*), and sea bass (*Dicentrarchus labrax*), which represent three key commercial species in the Bay of Biscay ecosystem. We specifically investigated the spatiotemporal variability in reproductive phenology at a monthly time step over eleven years between 2008 and 2018. Inferences derived from EOF analyses are compared to the literature to highlight the consistency of our results with regard

to the available knowledge of the phenology and location of reproduction. We also tested the influence of sea surface temperature on reproductive phenology as a posttreatment analysis.

## Materials and methods

### Outline of the approach

Our approach includes different steps, which are detailed hereafter:

- (i) Case studies and synthesis of the available knowledge on their reproductive phenology. Sole, hake, and sea bass are important fisheries of the Bay of Biscay. Based on a literature review, we provide expectations on the phenology and locations of reproductive areas.
- (ii) Inferring species distribution based on the ISTHM introduced by Alglave et al. (2022, 2023). We rely on the framework developed by Alglave et al. (2022, 2023) to map the biomass of the mature fraction of the population for each species (sole, hake, and sea bass) using a monthly time step from 2008 to 2018. The statistical approach integrates data from distinct trawler fleets that cover the whole Bay of Biscay.
- (iii) EOF and clustering analysis of the model outputs. To identify and visualize essential habitats and related seasons, we synthesized the seasonal (monthly) and interannual variation in the distribution of the abundance of mature fish using EOF analysis (realized independently for each species) followed by a clustering analysis.
- (iv) Investigating interannual variability of the reproduction phenology to environmental drivers. Finally, we interpret the main modes of variability of the EOF with regard to reproduction phenology. Our analysis allowed to investigate the spatiotemporal variability in reproduction phenology at two embedded time scales, the seasonal (intra-annual and monthly time steps) and interannual time scales.

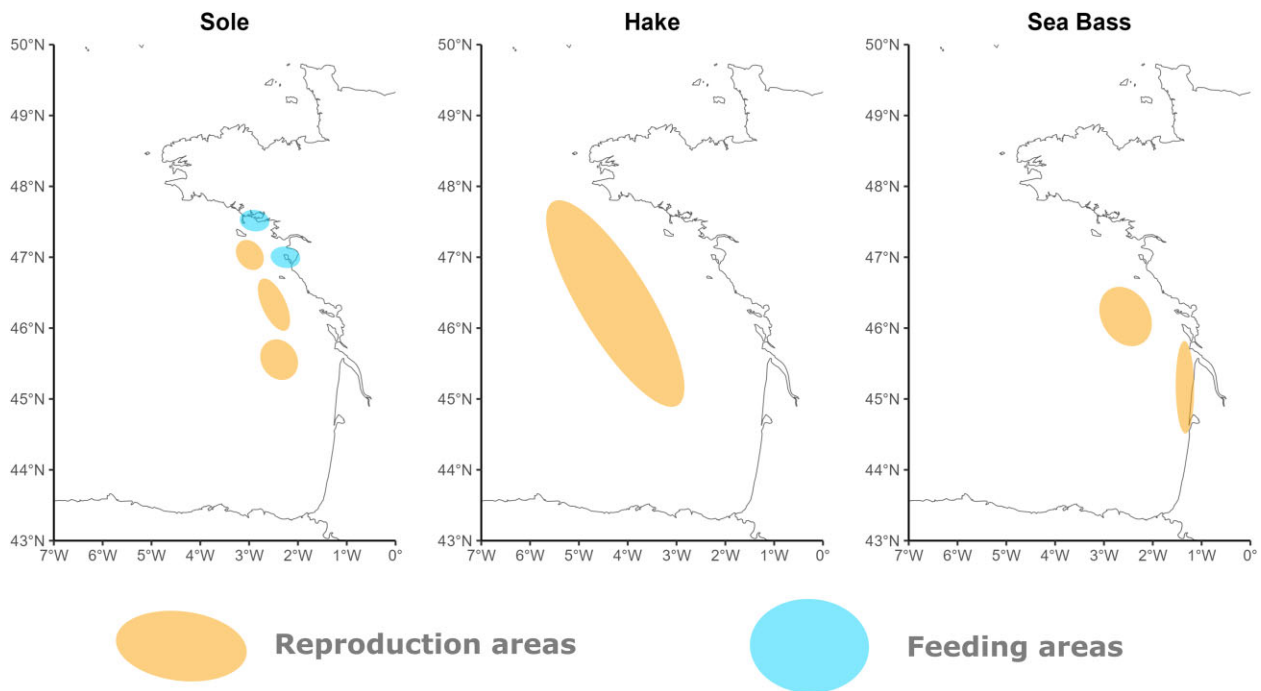
### Case studies

We selected three case studies of important species in the Bay of Biscay for which some information on essential habitats, especially on spawning grounds, is available but incomplete, including sole, hake, and sea bass (ICES 2020), as summarized in Fig. 1.

Arbault et al. (1986), Petitgas (1997), and Alglave et al. (2022) identified spawning grounds along the Bay of Biscay from January to March. Some feeding grounds are known along the Bay of Biscay coast (Koutsikopoulos et al. 1995; Fig. 1).

For hake, Alvarez et al. (2004) provided a similar analysis based on surveys conducted in the 1990s. Poulard (2001) investigated the rough-scale spatiotemporal distribution of hake based on logbook data. Woillez et al. (2007) developed spatial indices derived from annual survey data and provided a detailed description of hake spatial patterns over time; age 3 (~38 cm fish) appeared to be a turning point in these dynamics.

Recent analyses of seabass have investigated spawning area and timing based on "VMS  $\times$  logbooks" data and provided



**Figure 1.** Graphical synthesis of the available knowledge on essential habitats for the three species.

information on phenology (Dambrine et al. 2021). Reconstructed sea bass tracks obtained from large-scale tagging surveys confirmed sea bass to be a partial migratory species, as individuals exhibited either long-distance migration or residence. Most migrants exhibit fidelity to coastal summer feeding areas and offshore winter spawning areas, particularly on the central shelf of the Bay of Biscay (Pontual et al. 2019; 2023).

## Model structure and data to fit the model

### Data and commercial fleets

We analyzed the catch per unit of effort (CPUE) of trawlers (as in Alglave et al. 2022) over 11 years between 2008 and 2018, a relatively long period, to investigate interannual changes in species distribution and reproductive phenology.

As we only wanted to interpret the spatiotemporal dynamics of adult individuals, we filtered the mature fraction of the catch declarations by crossing catch declarations with the size distribution in each commercial category (see Alglave et al. 2023) for further details).

We selected data from several trawler fleets because they benefit from relatively opportunistic behavior and usually cover a wide area (Fig. 2). Furthermore, standard effort units can be relatively easily derived, and their CPUE provides a good indicator of the relative biomass of fish (Hovg rd and Lassen 2008). The selected fleets for each species are presented in Table 1.

### Model structure and spatiotemporal resolution

To map the spatiotemporal distribution of the mature fish biomass of these different species, we used the hierarchical integrated statistical framework developed by Alglave et al. (2022, 2023). The framework combines multiple data sources to infer the spatial distribution of fish biomass density. The model is fitted to the data between 2008 and 2018 at a

monthly time step (a total of  $11 \times 12 = 132$  time steps) on a  $0.05^\circ$  grid. It is structured in three layers:

- (i) The latent field of the spatial distribution of the mature biomass (the field we want to infer) defined at a monthly time step;
- (ii) The observation layer; this layer can handle CPUE data from different fleets, including the distinct catchability of the fleets; CPUE data are related to the same unique spatiotemporal field of relative abundance;
- (iii) Unknown parameters, including those that control the shape of the mature biomass latent field.

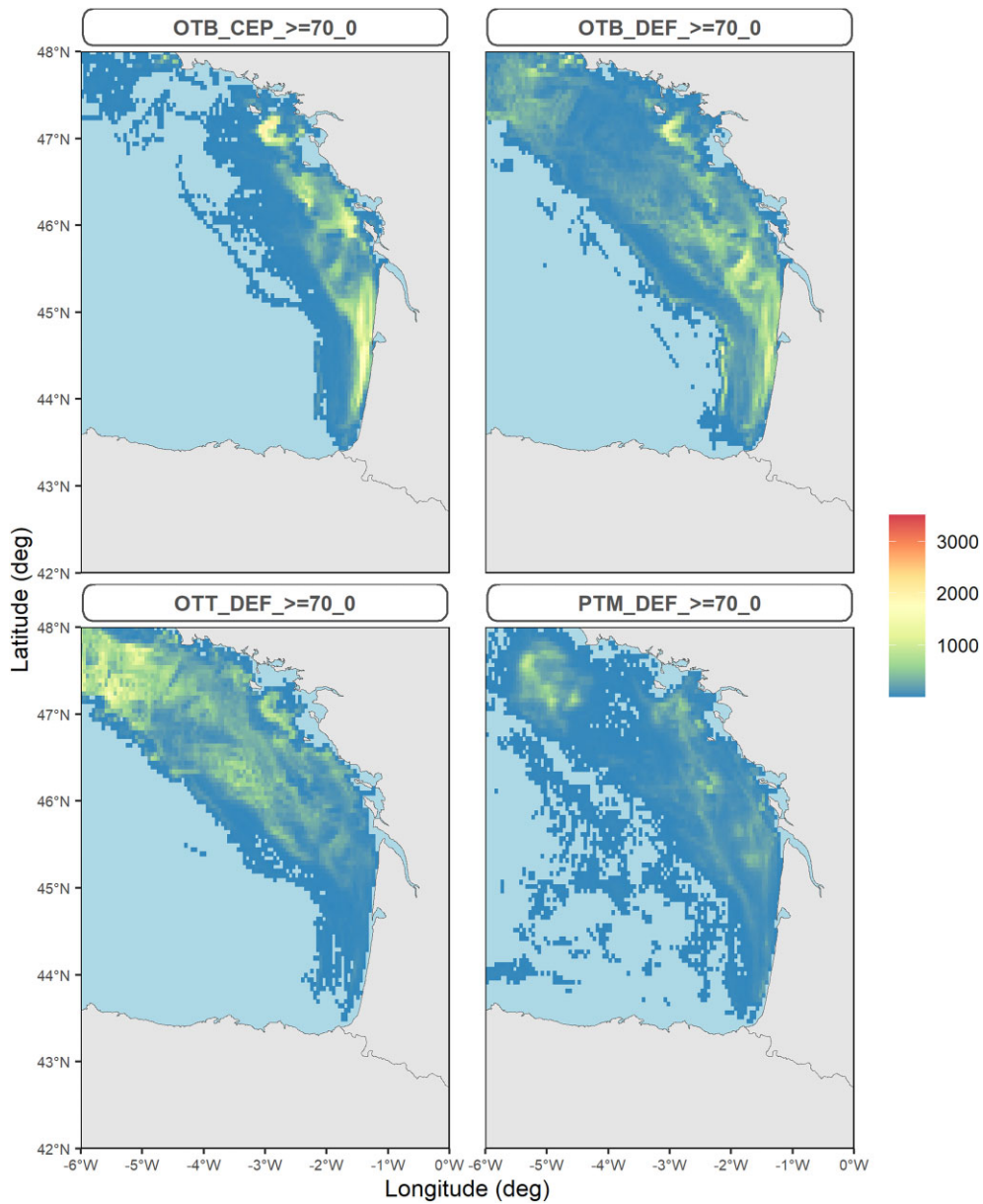
We simplified the framework developed by Alglave et al. (2022) by ignoring the preferential sampling of fishers. Indeed, previous results by Alglave et al. (2022) have shown that preferential sampling of trawlers is low. Taking this into account would therefore strongly increase the computational burden while only slightly affecting spatial predictions (Alglave et al. 2022).

## EOF to identify essential habitats and to highlight changes in phenology

### Basics of EOF

EOF was initially developed by Lorenz (1956) for weather forecasting. The broad idea is to generalize classical dimension reduction techniques such as principal component analysis to spatiotemporal dimensions. EOF seeks to summarize the information provided by a set of spatiotemporal maps into a smaller set of maps that best describe and summarize the spatiotemporal patterns of variability.

Let us define  $S(x, t)$  a biomass field defined at a time step  $t$  ( $t = \{1, \dots, T\}$ ), spatial cell  $x$ , and the centered field of biomass  $S^*(x, t) = S(x, t) - \overline{S(x, \cdot)}$  (with  $\overline{S(x, \cdot)}$  being the spatial average of  $S(x, t)$ ). Here,  $S^*(x, t)$  is expressed as a linear combination of spatial patterns  $p_m(x)$  (or maps, named EOF)



**Figure 2.** Spatial distribution of sampling effort (in hour) for each fleet aggregated over the period 2008–2018.

**Table 1.** Trawler fleets selected to infer species distribution for each species.

Species	Sole	Hake	Seabass
Fleets	OTB_DEF_≥70_0 bottom trawl targeting demersal fish OTB_CEP_≥70_0 bottom trawl targeting cephalopods OTT_DEF_≥70_0 otter trawl fleet targeting demersal	OTB_DEF_≥70_0 bottom trawl targeting demersal fish OTB_CEP_≥70_0 bottom trawl targeting cephalopods OTT_DEF_≥70_0 otter trawl fleet targeting demersal	OTB_DEF_≥70_0 bottom trawl targeting demersal fish OTB_CEP_≥70_0 bottom trawl targeting cephalopods PTM_DEF_≥70_0 pelagic trawl fleet targeting demersal fish

Note that the name of the fleets (e.g. OTB\_CEP\_≥70\_0) presents first the gear of the fleet (OTB for bottom trawl), then the species caught (CEP for cephalopods, DEF for demersal species) and the mesh size (≥70\_0 for 70 mm and above).

related to temporal indices (or loading factors)  $\alpha_m(t)$ .

$$S^*(x, t) = \sum_{m=1}^M \alpha_m(t) \times p_m(x); \quad x \in \{1, \dots, n\}, \\ t \in \{1, \dots, T\}, \quad M \leq T.$$

The loading factors  $\alpha_m(t)$  and the spatial patterns  $p_m(x)$  are defined to maximize the variation captured by the spatial patterns  $p_m(x)$  and to ensure that the spatial patterns and the loading factors are orthogonal. The first spatial map  $p_1(x)$  captures the greatest amount of spatial variation; the second spatial pattern  $p_2(x)$  is orthogonal to the first one and captures the second largest amount of spatial variation. In matrix terms, this falls back to a diagonalization problem and is equivalent to performing a PCA on a data frame where individuals are time steps and variables are locations (Lorenz 1956). Then,  $p_m(x)$  can also be called the eigen vectors and the classical PCA representation can be used to represent EOF results. Typically, the first two loading factors can be projected on the first two dimensions of the EOF to obtain a visual representation of the spatiotemporal decomposition of the signal on the first plane of variability. In practice, diagonalization is performed through singular value decomposition (Banerjee and Roy (2014)). It is available in R through the function *svd* (R Core Team 2023). Spatial patterns are normalized to 1, and loading factors are standardized using the square root of their eigenvalue.

### Hypotheses of linearity and stationarity of EOF

EOF relies on some hypotheses such as linearity and stationarity (Hannachi et al. 2007). Indeed, as in PCA, EOF accounts for linear correlations with the risk that other nonlinear relationships are not detected. Furthermore, although they are combined through loading factors that evolve over time, the spatial patterns [ $p_m(x)$ ] per se are stationary and will not change across the time series. In other words, at each time step, the distribution is modelled as a linear combination of the different spatiotemporal pattern (that are constant over time), but with a weight in the mixture that is defined by the loading factors (that vary over time). Then, the EOF method works like a mixture model with the loading factor that allows to capture temporal variations in the intensity of each spatial patterns in the mixture. The mixture can change over time (through the change in the loading factors), but the fundamental components of the mixture (the spatial patterns) remain constant. These are properties that help reducing complexity and allow for easier interpretability and implementation compared with more complex technics. As demonstrated with Vector Autoregressive Spatio-Temporal Model package (VAST), such hypotheses hold well for numerous fish species (Thorson et al. 2020a, 2020b). Still, they should be kept in mind, especially because any propagation patterns will not be evidenced through standard EOF. These properties are further detailed in the discussion.

### Filtering EOF dimensions and locations of the spatial pattern

For each species, we filter the number of dimensions based on the evolution of the (global) variance explained by adding new dimensions (scree plot). As a commonly used empirical rule of thumb, we cut the graph at the dimension where there is an angle in the graph of the variance explained (when there is no drop in variance explained in the next dimensions; Wike

et al. 2019). When plotting the spatial patterns, all the locations that contributed less than 1/(number of grid cells over the spatial domain) were shaded to highlight the locations that contributed most to the variation (local variance). In standard PCA, this is equivalent to keep only the variables (i.e. locations in our case) that explain or contribute more than a single variable (or location).

### Identifying seasons and essential habitats using clustering analysis

To identify distinct essential habitats and to relate these to the ecological season, two complementary clustering analyses were performed, one on the loading factors (temporal dimension) and the other on the EOF spatial maps (spatial dimension).

Following the analogy with PCA, in the EOF, individuals (here, time steps) and variables (here, locations) are projected into two distinct spaces: the space of the individuals (time steps) and the space of the variables (locations). Clustering can then be realized either in the space of individuals (time steps) or in the space of variables (locations). While clustering in the space of individuals seeks to identify groups of time steps that have similar spatial patterns, clustering in the space of variables seeks to identify locations with similar fish density at the same time step. Clusters of time steps will be interpreted as ecological seasons, and spatial clusters will be interpreted as distinct essential habitats. Clusters in the temporal and spatial dimensions can be represented in the first two dimensions of the EOF. Standardizing loading factors and the spatial EOF maps allows to plot them on the same scale and to compare the spatial clusters with the seasons identified through clustering. If loading factors (or seasons) are in the same direction as the locations (or spatial clusters), then they are correlated and can be associated (time steps  $\times$  locations) to identify the essential habitats and the period where individuals are in this habitat.

We performed clustering based on hierarchical clustering on principal components using the FactoMineR package (Lê et al. 2008).

By crossing the spatiotemporal patterns in the EOF with the literature review on reproductive habitat and phenology (Table 2; Fig. 1), it is possible to relate EOF principal components to reproductive phenological processes. Note that several dimensions could be related to a single process. For simplicity, we choose the dimension that best corresponds to the phenological process of interest (reproduction).

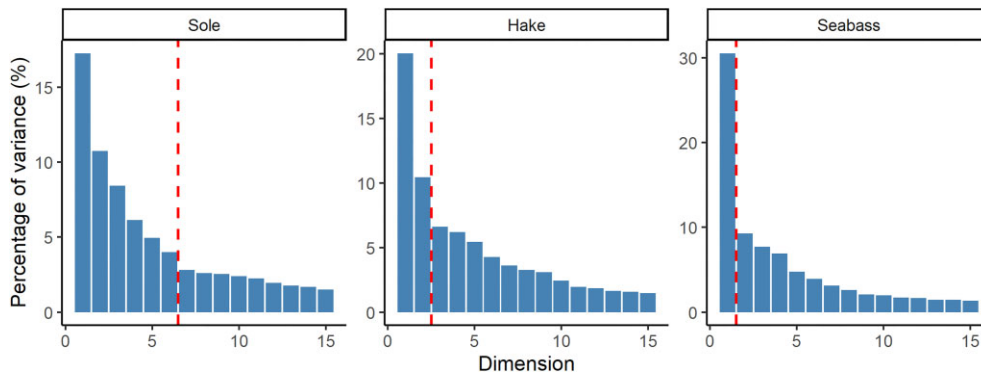
### Investigating interannual variability

Reproductive phenology is partly driven by temperature and can therefore vary between years (Fincham et al. 2013, Huret et al. 2018). Specifically, for sole, hake, and sea bass, previous studies have investigated the relationship between reproduction timing and sea surface temperature (SST) and have revealed an optimal range of temperatures for reproduction (references are provided in Table 2).

We hypothesized that the peak of the loading factor associated with the spawning season represents the peak of the reproduction season, and we investigated the interannual variability of the peak. We assessed whether the spawning peak identified from the EOF matched the optimal temperature range (e.g. see Fig. 4). SST data were extracted from the Marine Copernicus platform (<https://marine.copernicus.eu/>).

**Table 2.** Optimal range of temperature for each three species.

Species	Sole	Hake	Seabass
Optimal range temperature for reproduction	[10; 12.5°C]	[10; 12.5°C]	[11; 16°C]
Reference	Devauchelle (1986)	Murua (2010)	Devauchelle (1986)

**Figure 3.** Proportion of the total variance explained by each dimension of the EOF for each species. Dashed line: threshold used to filter the interpretable dimensions of the EOF.

## Results

### Extracting average spatial patterns from the EOF

The fitted biomass field  $S(x, t)$  averaged over months and anomalies  $S^*(x, t)$  are presented in Figs. S1.1, S1.2, and S1.3. For sole, we selected the first six EOF dimensions, which accounted for 50% of the variance (Fig. 3, see Fig. S3 for all six dimensions). For hake, we filter the first two dimensions, which capture 30% of the variance. For seabass, we selected only the first dimension, which captured 30% of the variance alone. Other dimensions are not considered in this analysis because they are considered as noise.

The averaged spatial distribution (denoted  $\overline{S(x, \cdot)}$  in the EOF equations) revealed specific average patterns for each case study (Fig. S2). For sole, the average distributions are relatively coastal with high biomass offshore of the Gironde Estuary (2°W–45°N). For hake, the average spatial distribution is greater offshore and corresponds to the slope area. For sea bass, the mean pattern is very coastal. Biomass is high along the Vendée coast (2°W–46°N to 3°W–47°N), with a hotspot near Belle Île (3°W–47°N), and along the Landes coast (1.5°W–44°N to 1.5°W–45.5°N).

### Combining EOF data with available knowledge to infer spawning phenology

All species present a strong seasonal pattern (Fig. 4). For sole, a periodic signal is revealed in the loading factors. EOF dimensions 1 and 2 highlight the high biomass in offshore areas in winter (December to April) and the relatively coastal distribution in summer. The first EOF spatial map (EOF1) mainly captures the coastal and offshore seasonal aggregations without highlighting spawning areas per se. In contrast, in the second dimension (EOF2), the offshore areas correspond to the reproduction grounds highlighted in Fig. 1. Additionally, the loading factors associated with EOF2 are maximal during the period when SSTs are favorable for reproduction. Therefore, for sole, EOF2 seems to be the best proxy for reproduction phenology. The orange areas in EOF2 are spawning areas, and the maximum loading factors are spawning peaks. Feeding areas

identified in the literature (Fig. 1) are very constrained in space while winter areas are more throughout the center of the Bay of Biscay (45°N to 47°N on the coast) with a strong anomaly off the Gironde estuary (2°W–45.5°N). These are considered as feeding areas.

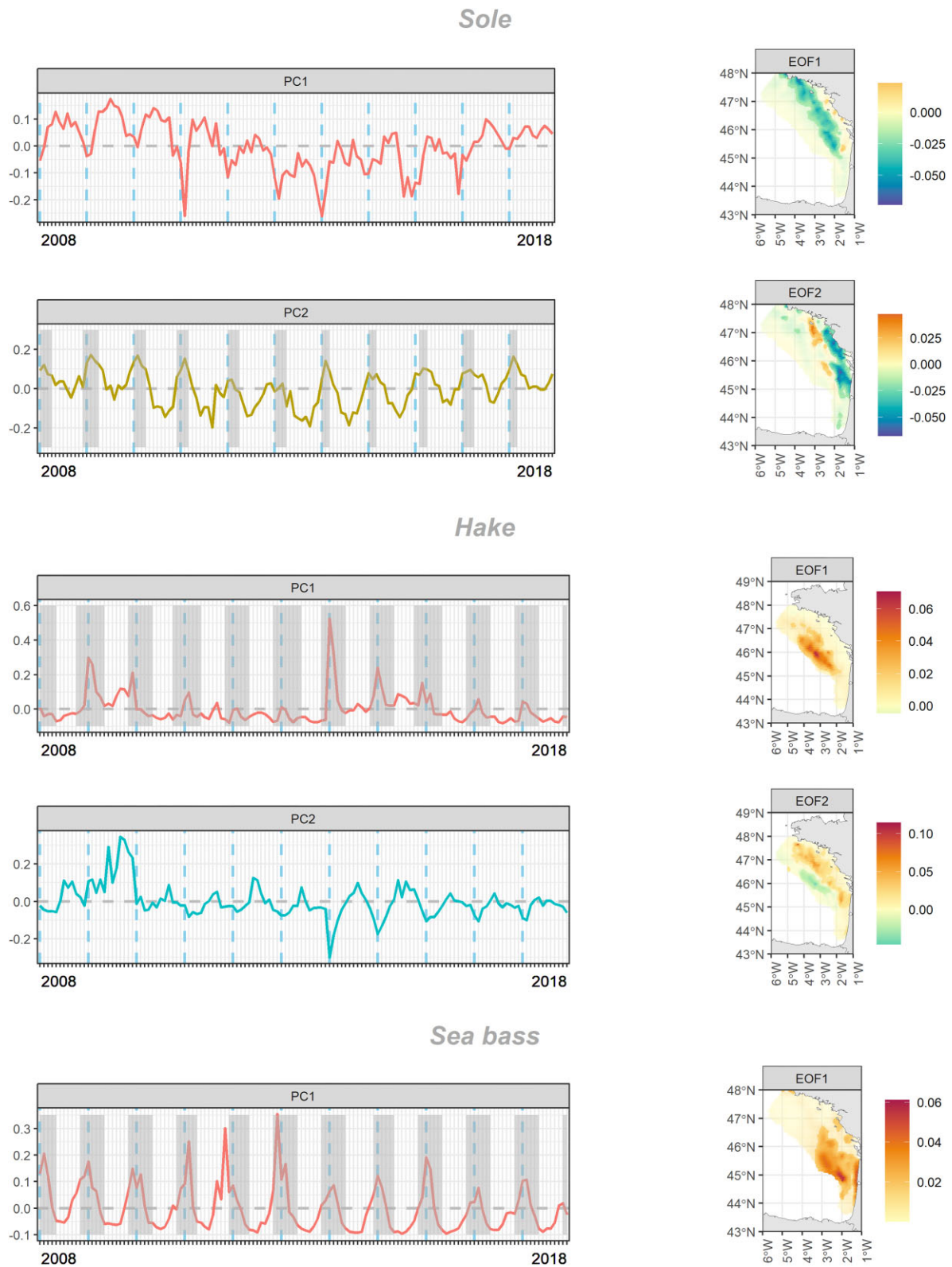
For hake, similar seasonal patterns are observed for dimensions 1 and 2. There are (1) shelf areas that are occupied during summer and (2) offshore areas on the edge of the shelf that are occupied during winter (3°W–46°N), which coincides with the spawning grounds shown in Fig. 1. The maximum loading factors associated with EOF1 occur within the period when the SST is favorable for reproduction. EOF2 represents offshore and coastal seasonal aggregations with less emphasis on reproduction. Hence, we retained EOF1 to investigate the phenology of hake reproduction.

For sea bass, EOF1 captures the aggregation in the central shelf of the Bay of Biscay, *i.e.* along the Landes coast (1.5°W–45°N) and outer reefs such as the “Plateau de Rochebonne” (2.5°W–46°N). One more area is evidenced in the EOF compared with prior literature knowledge (Fig. 1). This area is off the Gironde estuary (2°W–45°N) and can be considered as a reproduction area. The associated loading factors show a very strong seasonal pattern with high positive peaks occurring in January and February. These peaks match the periods during which the SST is favorable for reproduction and spawning, as shown in Fig. 1.

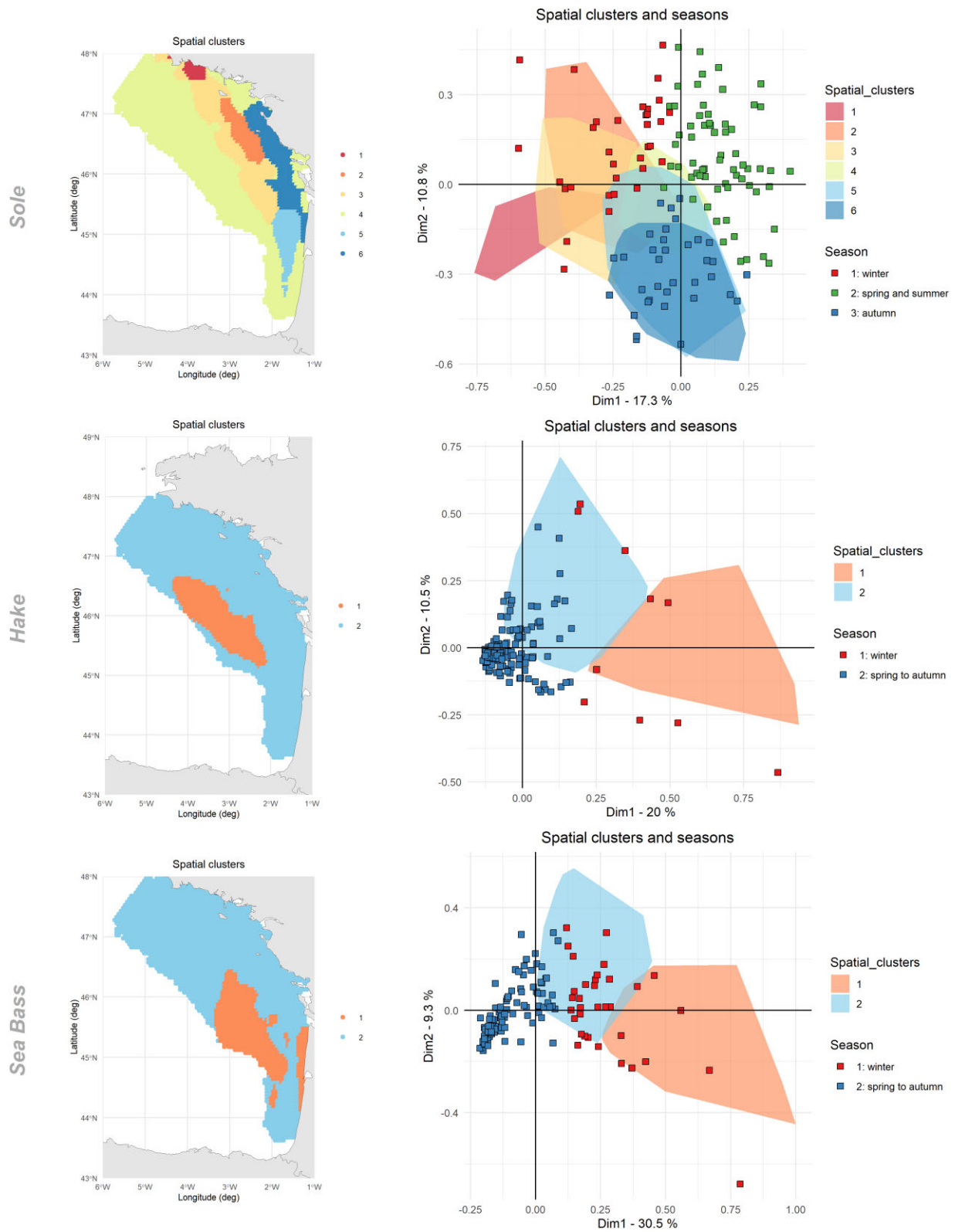
### Identifying the essential habitats and associated seasons

We perform two separate clustering analyses on the loading factors (time steps) and on the EOF spatial maps (eigenvectors). Combining the results allows to identify several combinations of areas and seasons that characterize spatiotemporal dynamics (Fig. 5).

For sole, six spatial clusters and three temporal clusters (three seasons) can be identified (Fig. 5; all the clustering trees are available in Fig. S5.1–S.6):

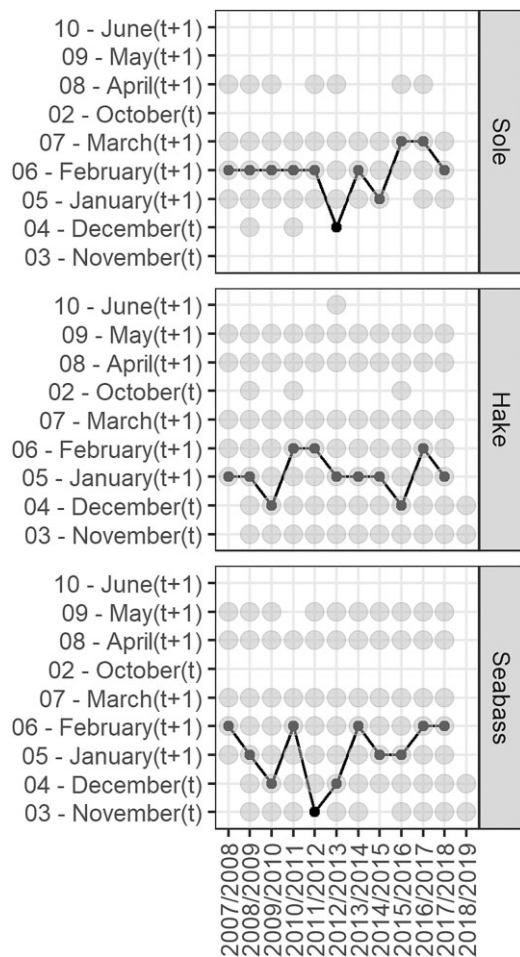


**Figure 4.** Loading factors (left) and EOF maps (right) for each species. For sole, only the two first dimensions are presented, the other dimensions are presented in Fig. S2. The dashed vertical line is the month of January for each year. The dashed horizontal line is the 0 value. The shaded bands are the period when temperature is within the optimal range of temperature for reproduction. The shaded points of the spatial maps represent the areas that do not contribute strongly to the variance. The local variance explained by the two first EOF dimensions for each species is available in Fig. S4.



**Figure 5.** Essential habitats and associated seasons. (Left column) Mapping spatial clusters using eigenvectors to identify essential habitat. (Right column) Projection of the standardized loading factors (seasons) and the standardized eigen vectors (spatial clusters) on the two first dimensions of the EOF. The same colors are used to represent spatial cluster on the maps (left column) and on the first two dimensions of the EOF (polygons in the right-hand columns). On the right-hand column, each season is represented by a square of a specific color. If the loadings of a season are in the same direction as a spatial cluster, then they are correlated and they correspond to a couple (season × essential habitat). The time steps are displayed in Fig. S6.





**Figure 6.** Variability of the period of reproduction identified based on the peak in the time series of loading factors associated with the selected EOF dimensions (Fig. 4). For sole, we consider the second dimension of the EOF. For hake, we consider the minimum first dimension of the EOF. For sea bass, we consider the maximum of the first dimension of the EOF. The gray dots are the months for which the average temperature is within the optimal range of temperature for reproduction.

- (1) an area constituted by Clusters 1–3 that is mainly correlated with winter months and that is considered as a reproduction area. This is consistent with previous knowledge of essential habitats (Fig. 1);
- (2) an area constituted by Cluster 4 that mainly corresponds to the average distribution for spring and summer;
- (3) a coastal area constituted by Clusters 5 and 6 that mainly correlates with end summer and autumn months that corresponds to feeding ground.

Note that the PCA representation of the loading factors and seasons with the related time steps are in Fig. S6.

For hake and sea bass, a similar workflow was used to identify essential habitats. Hake aggregate during winter in offshore areas and return to the average distribution during the other months (Fig. 5, centerline). This is consistent with the available knowledge synthesized in Fig. 1. Sea bass aggregate all along the Landes coast and on the shelf of the Bay of Biscay during winter (Fig. 5; bottom line). Specifically, the area off the Gironde estuary (2°W–45°N) is an additional reproduction area that is not referenced in the literature

(Fig. 1). It returns to an average distribution on the other time steps.

### Interannual variability in reproduction and its relationship with SST

Our results also highlight the interannual variability in reproductive phenology for the three species (Fig. 6). For sole, the months of reproduction identified through EOF2 fall between January and March. In 2012–2013, reproduction seems to have occurred slightly earlier. The maximum loading factors are obtained in December and occur outside the period where temperature is favorable for reproduction. During the 2012–2013 period (Fig. 4, PC2 for sole), loading factors exhibited lower temporal variability than in the other years, and reproduction could also occur later (there was another peak in March).

For hake, the months of reproduction are slightly earlier, namely, between December and February.

For sole and hake, both the reproduction period and the time range where temperature is favorable appear to be relatively stable.

For sea bass, reproduction months emphasize more variability. The maximum loading factors associated with EOF1 occur between November and February, specifically at the beginning of the time series. In contrast, the period during which the SST is favorable for reproduction is steady. This suggests that covariates other than temperature may strongly affect reproductive timing.

## Discussion

In this paper, we combined an integrated species distribution model informed by “VMS × logbooks” catch declaration data with a dimension-reduction approach (EOF) to investigate the reproductive phenology of three key species of the Bay of Biscay ecosystem and fisheries (sole, hake, and sea bass). “VMS × logbooks” data open new gates for ecological analysis at a much finer spatiotemporal resolution than ever before (Gerritsen and Lordan 2011, Murray et al. 2013, Azevedo and Silva 2020). To our best knowledge, our study is among the first ones that demonstrated this potential through concrete analysis of large time periods with massive amounts of “VMS × logbooks” data.

Combining logbook catch declarations with VMS data provides reliable catch per unit effort of mature fish, which were used to infer distributions of mature fish biomass density with a monthly time step over the 2008–2018 period. Posttreatment of spatiotemporal distribution predictions through EOF methods provides a method to analyze reproduction phenology at a much finer temporal scale than other data sources (e.g. scientific surveys).

In the response to the three points addressed in the introduction, for all three case studies, (i) we identified spatio-seasonal patterns that match the reproduction phenology. We identified essential habitats that were already known as well as some that were poorly known or not yet identified, such as feeding grounds (here for hake and for sole). (ii) We investigated the interannual variations in the reproductive ecology of each species by constructing a phenological index and (iii) we related this to a key covariate of reproduction phenology here temperature. Such modeling approach support

broader access to VMS data and their analysis to improve ecological knowledge (Hintzen *et al.* 2012).

### The need to combine statistical results with the literature to interpret the results

EOF captures the main modes of variation of the spatiotemporal fields. The identification of spatiotemporal patterns through EOF is not self-sufficient, and its interpretation in terms of reproduction phenology relies partly on the availability of expert and/or literature knowledge.

Typically, for sole, some studies investigating the timing and spatial distribution of egg and larval distributions have been used to corroborate our interpretation of EOF (Arbault *et al.* 1986, Petitgas 1997). For hake, survey data have recorded mature individuals during the spawning season (Alvarez *et al.* 2004), and catch declaration data have been used to evidence reproduction migration at the level of rough statistical rectangles (Poulard 2001). Woillez *et al.* (2007) provided a detailed description of hake spatial patterns, with older ages being further offshore and aligning with the continental shelf. These findings were also consistent with our analysis.

Nevertheless, these data are limited. In both cases, the data were restricted to a few samples from very old spring surveys, and they have not been updated since this period. Furthermore, in many other cases, there might not be any reference data to corroborate the EOF results. We recommend conducting additional surveys, such as egg surveys, to obtain direct observations of spawning areas (Fox *et al.* 2008).

Alternatively, expert knowledge of fishers could be a valuable complement to interpret the main modes of variation (Yochum *et al.* 2011) when other data are missing. Bezerra *et al.* (2021) and Silvano *et al.* (2006) demonstrated the usefulness of fishers' knowledge for determining the temporality of fish spawning and identifying spawning grounds by combining information on aggregation areas provided by several fishers.

### An enhanced EOF method to better disentangle ecological processes

An extensive literature on EOF has outlined both its usefulness and efficiency for dimensionality reduction and its limitations (Hannachi *et al.* 2007). Several authors have indicated that extra care needs to be taken when interpreting statistical structure as a physical or ecological process (Monahan *et al.* 2009). Typically, a single process can be reflected through several modes of the EOF. In our application, we decided to restrict the interpretation to the one or two dimensions that best highlight reproductive phenology. Nevertheless, some reproduction signals can be observed in the other dimensions of the EOF. Other methods (in general, linear transformations of EOF) have been developed to better disentangle the underlying ecological process (Hannachi *et al.* 2007). For instance, a rotated EOF allows more distinct patterns to be obtained by relaxing orthogonality constraints. However, the choice of the rotation criterion and the number of dimensions used for rotation have to be set arbitrarily, which implies the need to make additional choices. Also, we applied EOF as a post-treatment on fitted values from a first species distribution model, and our method does not inherently account for uncertainties, which may be critical in fisheries, where data quality and uncertainty are common challenges. Further research is needed

to derive guidelines for best practices for such techniques to derive ecological inferences.

### The hypotheses of linearity and stationarity underlying EOF

Another limitation of this method is the linearity and stationarity of the EOF decomposition as discussed previously (Cressie and Wikle 2015). In marine ecological data, especially in a context of climate change, the underlying processes can be highly nonlinear and nonstationary (Ciannelli *et al.* 2012). Hence, EOF may not capture important nonlinear relationships or complex spatial patterns that exist in the data. These methods may oversimplify the spatiotemporal dynamics of fish populations. The development of spatiotemporal dimension reduction methods that can account for nonstationarity or nonlinearity is an open research avenue; it would be of great interest to help disentangle ecological processes while accounting for the uncertainty contained in the data.

### Limits of using commercial catch data

Another limitation of our approach is that the use of commercial catch declarations hampers the ability to analyze phenology relative to that of smaller size classes. In our approach, we filtered the mature component of the populations based on the size distribution of each commercial category as described in Alglave *et al.* (2023). A similar approach was used by Azevedo and Silva (2020) (although their approach was more refined, as proportion by length was considered to vary by size category and by zone) and allowed the mapping of different age classes of horse mackerel. However, because the fishery is size selective, a strong limitation of our approach is that commercial catch declaration data likely provide a biased picture of the spatiotemporal distribution of smaller fish size classes. The minimum landing size is often relatively high, and individuals below this size are either rejected or not declared (Lehuta and Vermard 2023). Furthermore, VMS data are not available for vessels below 12 m, which may represent an important part of coastal vessel fishing near juvenile habitats.

As a consequence, mapping juveniles' habitat is not possible when exclusively using catch declaration data. In this case, nursery surveys remain the only available data for investigating juvenile spatiotemporal dynamics and should be considered a reference (e.g. Nurse, Nourseine et Noursom—Delaunay and Brind'Amour 2018 or Nourdem survey series—Roy *et al.* 2022).

Another limitation when using commercial data is the targeting behavior of fishers. This can lead to biased spatial predictions and to overestimated estimates of biomass. Here, we neglected this because trawlers do not exhibit strong targeting behavior. Nevertheless, the framework developed by Alglave *et al.* (2022) could be useful in the case of stronger preferential sampling (see Quemper 2021).

### Investigating the effect of environmental drivers on phenology

Temperature is an important factor in fish reproduction (Huret *et al.* 2018). Other covariates may also strongly affect the timing of reproduction. This is especially true for species such as sea bass, which spend part of their life cycle in the pelagic realm. Other factors, such as salinity or the concentration of chlorophyll A or light, could affect reproductive timing

(Vinagre et al. 2009, Planque et al. 2011). The literature does not explicitly mention any clear threshold for these covariates; thus, more extensive field sampling would be required to identify the determinants of reproduction and include these determinants in our approach.

Furthermore, it is expected that in a changing environment, the duration of reproduction will shift with temperature. Some results already illustrate such a phenomenon for sole in the North Sea, the English Sea, and the North East Channel (Fincham et al. 2013). We consider our approach is interesting for assessing the effect of climate change on fish phenology to preserve fish habitats and ensure the renewal of marine resources in the context of rapid environmental changes.

## Acknowledgement

The authors are grateful to the Direction Générale des Affaires Maritimes, de la Pêche et de l'Aquaculture (DGAMPA) and Ifremer (Système d'Informations Halieutiques—SIH) who provided the aggregated VMS data. The findings and conclusions of the present paper are those of the authors.

## Author contributions

All authors contributed to the conceptualization and methodology of the study. All authors contributed to analysis of findings as well as drafting and revising the manuscript.

## Supplementary data

[Supplementary data](#) is available at *ICES Journal of Marine Science* online.

*Conflict of interest:* The authors have no conflicts of interest to declare.

## Funding

We thank the Direction générale de l'énergie et du climat (DGEC), who, within the framework of “L'Observatoire de l'Eolien en Mer”, funded the time of Juliette Casemajor and Mathieu Woillez, the packaging of R codes used in this study (R package “FishMap”) and the open access of this article.

## Data availability

Logbooks and VMS data are confidential data and they are available on specific request to DPMA. Model outputs for the different species are available on zenodo (<https://zenodo.org/records/10517096>). Codes are available on Github (<https://github.com/balglave/C4-EOF.git>).

## References

- Alglave B, Rivot E, Etienne MP *et al.* Combining scientific survey and commercial catch data to map fish distribution. *ICES J Mar Sci* 2022;79:1133–49. <https://doi.org/10.1093/icesjms/fsac032>
- Alglave B, Vermard Y, Rivot E *et al.* Identifying mature fish aggregation areas during spawning season by combining catch declarations and scientific survey data. *Can J Fish Aquat Sci* 2023;80:808–24. <https://doi.org/10.1139/cjfas-2022-0110>
- Alvarez P, Fives J, Motos L *et al.* Distribution and abundance of European hake *Merluccius merluccius* (L.), eggs and larvae in the North East Atlantic waters in 1995 and 1998 in relation to hydrographic conditions. *J Plankton Res* 2004;26:811–26. <https://doi.org/10.1093/plankt/fbh074>
- Arbault PS, Camus P, le Bec C. Estimation du stock de sole (*Solea vulgaris*, Quensel 1806) dans le Golfe de Gascogne à partir de la production d'œufs. *J Appl Ichthyol* 1986;2:145–56. <https://doi.org/10.1111/j.1439-0426.1986.tb00656.x>
- Azevedo M, Silva C. 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 J Mar Sci* 2020;77:2933–44. <https://doi.org/10.1093/icesjms/fsaa170>
- Banerjee S, Roy A. *Linear Algebra and Matrix Analysis for Statistics*, 1st edn. Boca Raton Fla: Chapman and Hall/CRC, 2014.
- Bastardie F, Nielsen JR, Ulrich C *et al.* Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. *Fish Res* 2010;106:41–53. <https://doi.org/10.1016/j.fishres.2010.06.016>
- Bezerra IM, Hostim-Silva M, Teixeira JLS *et al.* 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. *Fish Res* 2021;239:105937. <https://doi.org/10.1016/j.fishres.2021.105937>
- Biggs CR, Heyman WD, Farmer NA *et al.* The importance of spawning behavior in understanding the vulnerability of exploited marine fishes in the US Gulf of Mexico. *PeerJ* 2021;9:e11814. <https://doi.org/10.7717/peerj.11814>
- Ciannelli L, Bartolino V, Chan KS. Non-additive and non-stationary properties in the spatial distribution of a large marine fish population. *Proc R Soc B Biol Sci* 2012;279:3635–42. <https://doi.org/10.1098/rspb.2012.0849>
- Cressie N, Wikle CK. *Statistics for spatio-temporal data*. Hoboken, New Jersey: John Wiley & Sons, 2015.
- Dambrine C, Woillez M, Huret M *et al.* Characterising essential fish habitat using spatio-temporal analysis of fishery data: a case study of the European seabass spawning areas. *Fisher Oceanogr* 2021;30:413–28. <https://doi.org/10.1111/fog.12527>
- de Pontual H, Lalire M, Fablet R *et al.* New insights into behavioural ecology of European seabass off the West Coast of France: implications at local and population scales. *ICES J Mar Sci* 2019;76:501–15. <https://doi.org/10.1093/icesjms/tsy086>
- de Pontual H, Heerah K, Goossens J *et al.* Seasonal migration, site fidelity, and population structure of European seabass (*Dicentrarchus labrax*). *ICES J Mar Sci* 2023;80:1606–18. <https://doi.org/10.1093/icesjms/fsad087>
- Delage N, Le Pape O. *Inventaire des zones fonctionnelles pour les ressources halieutiques dans les eaux sous souveraineté française. Première partie: Définitions, critères d'importance et méthode pour déterminer des zones d'importance à protéger en priorité*. Rennes: Pôle halieutique AGROCAMPUS OUEST, 2016.
- Delaunay D, Brind'Amour A. Campagnes « Nourriceries ». 2018. <https://archimer.ifremer.fr/doc/00435/54610/>
- Devauchelle N. Ponte en captivité et incubation de quatre poissons marins élevés en Europe. In: *Conférence MEDRAP (FAO): Production contrôlée d'alevins en Europe-Zadar/Pula (Yougoslavie, 10-28 février 1986)*, 1986.
- Fincham JI, Rijnsdorp AD, Engelhard GH. Shifts in the timing of spawning in sole linked to warming sea temperatures. *J Sea Res* 2013;75:69–76. <https://doi.org/10.1016/j.seares.2012.07.004>
- Fox CJ, Taylor M, Dickey-Collas M *et al.* Mapping the spawning grounds of North Sea cod (*Gadus morhua*) by direct and indirect means. *Proc R Soc B Biol Sci* 2008;275:1543–8. <https://doi.org/10.1098/rspb.2008.0201>
- Gerritsen H, Lordan C. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J Mar Sci* 2011;68:245–52. <https://doi.org/10.1093/icesjms/fsq137>
- Grüss A, Biggs CR, Heyman WD *et al.* Protecting juveniles, spawners or both: a practical statistical modelling approach for the design of

- marine protected areas. *J Appl Ecol* 2019;56:2328–39. <https://doi.org/10.1111/1365-2664.13468>
- Grüss A, Thorson JT, Stawitz CC et al. Synthesis of interannual variability in spatial demographic processes supports the strong influence of cold-pool extent on eastern Bering Sea walleye pollock (*Gadus chalcogrammus*). *Prog Oceanogr* 2021;194:102569. <https://doi.org/10.1016/j.pocean.2021.102569>
- Hannachi A, Jolliffe IT, Stephenson DB. Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int J Climatol* 2007;27:1119–52. <https://doi.org/10.1002/joc.1499>
- Harden J. Fish Migration. *Copeia* 1969;2:409–11.
- Hintzen NT, Bastardie F, Beare D et al. VMStools: Open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data. *Fish Res* 2012;115–116:31–43. <https://doi.org/10.1016/j.fishres.2011.11.007>
- Hovgård H, Lassen H. *Manual on estimation of selectivity for gillnet and longline gears in abundance surveys*. Food & Agriculture Org, 2008, 397.
- Huret M, Bourriau P, Doray M et al. Survey timing vs. ecosystem scheduling: Degree-days to underpin observed interannual variability in marine ecosystems. *Prog Oceanogr* 2018;166:30–40. <https://doi.org/10.1016/j.pocean.2017.07.007>
- ICES. Working Group for the Bay of Biscay and the Iberian waters Ecoregion (WGBIE). *ICES Scientific Reports* 2020;2:845.
- Jolly GM. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* 1965;52:225–47. <https://doi.org/10.2307/2333826>
- Kai M, Thorson JT, Piner KR et al. Spatiotemporal variation in size-structured populations using fishery data: An application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Can J Fish Aquat Sci* 2017;74:1765–80. <https://doi.org/10.1139/cjfas-2016-0327>
- Koutsikopoulos C, Dorel D, Desauany Y. Movement of sole (*Solea solea*) in the Bay of Biscay: coastal environment and spawning migration. *J Mar Biol Assoc UK* 1995;75:109–26.
- Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw* 2008;25:1–18. <https://doi.org/10.18637/jss.v025.i01>
- Lehuta S, Vermard Y. Contrasting impacts of the landing obligation at fleet scale: impact assessment of mitigation scenarios in the Eastern English Channel. *ICES J Mar Sci* 2023;80:518–31. <https://doi.org/10.1093/icesjms/fsac148>
- Lieth H. *Phenology and Seasonality Modeling*. New York: Springer Science & Business Media, 2013.
- Lorenz EN. *Empirical orthogonal functions and statistical weather prediction*. Cambridge, Massachusetts: Massachusetts Institute of Technology, Department of Meteorology, 1956.
- Monahan AH, Fyfe JC, Ambaum MHP et al. Empirical orthogonal functions: the medium is the message. *J Clim* 2009;22:6501–14. <https://doi.org/10.1175/2009JCLI3062.1>
- Murray LG, Hinz H, Hold N et al. The effectiveness of using CPUE data derived from Vessel Monitoring Systems and fisheries logbooks to estimate scallop biomass. *ICES J Mar Sci* 2013;70:1330–40. <https://doi.org/10.1093/icesjms/fst099>
- Murua H. The biology and fisheries of European hake, *Merluccius merluccius*, the north-east Atlantic. *Adv Mar Biol* 2010;58:97–154.
- Olmos M, Ianelli J, Ciannelli L et al. Estimating climate-driven phenology shifts and survey availability using fishery-dependent data. *Prog Oceanogr* 2023;215:103035.
- Petitgas P. Sole egg distributions in space and time characterised by a geostatistical model and its estimation variance. *ICES J Mar Sci* 1997;54:213–25.
- Petitgas P, Doray M, Huret M et al. Modelling the variability in fish spatial distributions over time with empirical orthogonal functions: anchovy in the Bay of Biscay. *ICES J Mar Sci* 2014;71:2379–89. <https://doi.org/10.1093/icesjms/fsu111>
- Planque B, Loots C, Petitgas P et al. Understanding what controls the spatial distribution of fish populations using a multi-model approach. *Fisher Oceanogr* 2011;20:1–17. <https://doi.org/10.1111/j.1365-2419.2010.00546.x>
- Poulard J-C. Distribution of hake (*Merluccius merluccius*, Linnaeus, 1758) in the Bay of Biscay and the Celtic sea from the analysis of French commercial data. *Fish Res* 2001;50:173–87.
- Quemper F. *Modélisation de la distribution spatiale de la sardine du Golfe de Gascogne (*Sardina pilchardus*) par intégration de données commerciales et scientifiques: enjeux et limites*. Rennes: Institut Agro, 2021.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2023.
- Randon M, Réveillac E, Le Pape O. A holistic investigation of tracers at population and individual scales reveals population structure for the common sole of the Eastern English Channel. *Estuar Coast Shelf Sci* 2021;249:107096.
- Roy A, Lebigre C, Drogou M et al. Estimating abundance indices of juvenile fish in estuaries using Geostatistics: an example of European sea bass (*Dicentrarchus labrax*). *Estuarine Coastal Shelf Sci* 2022;269:107799. <https://doi.org/10.1016/j.ecss.2022.107799>
- Rufener M-C, Kristensen K, Nielsen JR et al. Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. *Ecol Appl* 2021;31:e02453. <https://doi.org/10.1002/eap.2453>
- Seber GA. A note on the multiple-recapture census. *Biometrika* 1965;52:249–60. <https://doi.org/10.1093/biomet/52.1-2.249>
- Silvano RA, MacCord PF, Lima RV et al. When does this fish spawn? Fishermen's local knowledge of migration and reproduction of Brazilian coastal fishes. *Environ Biol Fishes* 2006;76:371–86. <https://doi.org/10.1007/s10641-006-9043-2>
- Thorson JT, Cheng W, Hermann AJ et al. Empirical orthogonal function regression: Linking population biology to spatial varying environmental conditions using climate projections. *Global Change Biol* 2020a;26:4638–49. <https://doi.org/10.1111/gcb.15149>
- Thorson JT, Ciannelli L, Litzow MA. Defining indices of ecosystem variability using biological samples of fish communities: A generalization of empirical orthogonal functions. *Prog Oceanogr* 2020b;181:102244. <https://doi.org/10.1016/j.pocean.2019.102244>
- Vinagre C, Ferreira T, Matos L et al. Latitudinal gradients in growth and spawning of sea bass, *Dicentrarchus labrax*, and their relationship with temperature and photoperiod. *Estuarine Coastal Shelf Sci* 2009;81:375–80. <https://doi.org/10.1016/j.ecss.2008.11.015>
- Wikle CK, Zammit-Mangion A, Cressie N. *Spatio-temporal statistics with R*. Boca Raton, Florida: Chapman and Hall/CRC, 2019.
- Wuillez M, Poulard J-C, Rivoirard J et al. Indices for capturing spatial patterns and their evolution in time, with application to European hake (*Merluccius merluccius*) in the Bay of Biscay. *ICES J Mar Sci* 2007;64:537–50. <https://doi.org/10.1093/icesjms/fsm025>
- Wuillez M, Petitgas P, Huret M et al. Statistical monitoring of spatial patterns of environmental indices for integrated ecosystem assessment: application to the Bay of Biscay pelagic zone. *Prog Oceanogr* 2010;87:83–93. <https://doi.org/10.1016/j.pocean.2010.09.009>
- Yochum N, Starr RM, Wendt DE. Utilizing fisherman knowledge and expertise: keys to success for collaborative fisheries research. *Fisheries* 2011;36:593–605. <https://doi.org/10.1080/03632415.2011.633467>

Handling Editor: Sam Subbey