ICES Journal of Marine Science

ICES Journal of Marine Science (2014), 71(9), 2379–2389. doi:10.1093/icesjms/fsu111

Original Article

Modelling the variability in fish spatial distributions over time with empirical orthogonal functions: anchovy in the Bay of Biscay

Pierre Petitgas[1]*, Mathieu Doray[1], Martin Huret[2], Jacques Massé[1], and Mathieu Woillez[2]

[1]*IFREMER, Unit EMH, BP 22105, 44300 Nantes, France*
[2]*IFREMER, Unit STH, BP 70, 29280 Plouzané, France*

**Corresponding author: tel: +33 240 37 41 63; e-mail: pierre.petitgas@ifremer.fr*

Petitgas, P., Doray, M., Huret, M., Massé, J., and Woillez, M. Modelling the variability in fish spatial distributions over time with empirical orthogonal functions: anchovy in the Bay of Biscay. – ICES Journal of Marine Science, 71: 2379–2389.

Received 25 April 2014; revised 25 May 2014; accepted 27 May 2014; advance access publication 24 June 2014.

Characterizing the space–time variability in spatial distributions as well as understanding its drivers is basic to designing robust spatial management plans. As a prerequisite, we analyse here how this variability relates to population dynamics in conjunction with environmental conditions. For that, spatio-temporal statistical approaches are needed but seldom used in fisheries science. To fill this gap, we showcase the usefulness of the method of empirical orthogonal functions (EOFs). Guidelines are given to apply the method on a series of gridded maps as derived from fisheries survey data-series that now span over decades. The method is applied to the series, 2000–2012, of the spatial distributions of European anchovy in the Bay of Biscay at spawning time. Across the series, the EOF decomposition allowed to identify three main types of spatial distributions. One type corresponded to an extended distribution, another to a restricted distribution in core areas, and the third to a very coastal distribution. The coastal spawning distribution corresponded to a low population growth rate as it was never followed by a large recruitment in the subsequent year. We did not attempt to explain the spatial patterns *per se* but the drivers of change from one type of distribution to another. Stock size and fish size as well as bottom temperature and water column stratification were the covariates that controlled the variability in the spatial distributions over time. Further, the spatial distribution at spawning time related to recruitment in the following year, meaning that variability in the spatial distribution of spawning affected population dynamics. The typology of maps based on EOF decomposition summarized this spatial variability into spatial spawning configurations, which may serve spatial planning.

Keywords: anchovy, Bay of Biscay, density-dependence, EOF, habitat, spatial distribution.

Introduction

Habitats represent the environmental conditions that are favourable for an organism (e.g. for its presence, growth) and thus habitat maps provide the space–time envelops of suitable conditions. Statistical regression has been widely used to model the habitats of the presence of species (Guisan and Zimmermann, 2000; Austin, 2007) or fish populations (Planque *et al.*, 2011; Le Pape *et al.*, 2014). But although closely related, habitats and spatial distributions are different concepts. Even if habitats are potentially suitable, their occupation will rely on the ability of the fish to colonize them with varying density. Thus, to link habitats (suitable conditions) to fish spatial distributions, one needs to consider the mechanisms of habitat occupancy. These mechanisms involve factors internal to the population (e.g. abundance, demography, behaviour) as well as interactions in the ecosystem (e.g. trophic interactions, connectivity across the life cycle). We shall here focus on the former factors. Density-dependent habitat selection models have been used to explain how population spatial distributions vary with overall population abundance in different ways (MacCall, 1990; Shepherd and Litvak, 2004). Also physiological and behavioural mechanisms have been invoked to explain the re-colonization of past habitats during the rebuilding phase of a stock after its collapse (Petitgas *et al.*, 2010). Thus, we shall here consider that variability in spatial distributions results from variability in environmental conditions and internal population behaviour (Figure 1).

Further, we hypothesize that the spatial distribution of a population at spawning time is not independent of its demographic dynamics as it affects subsequent recruitment and therefore there is

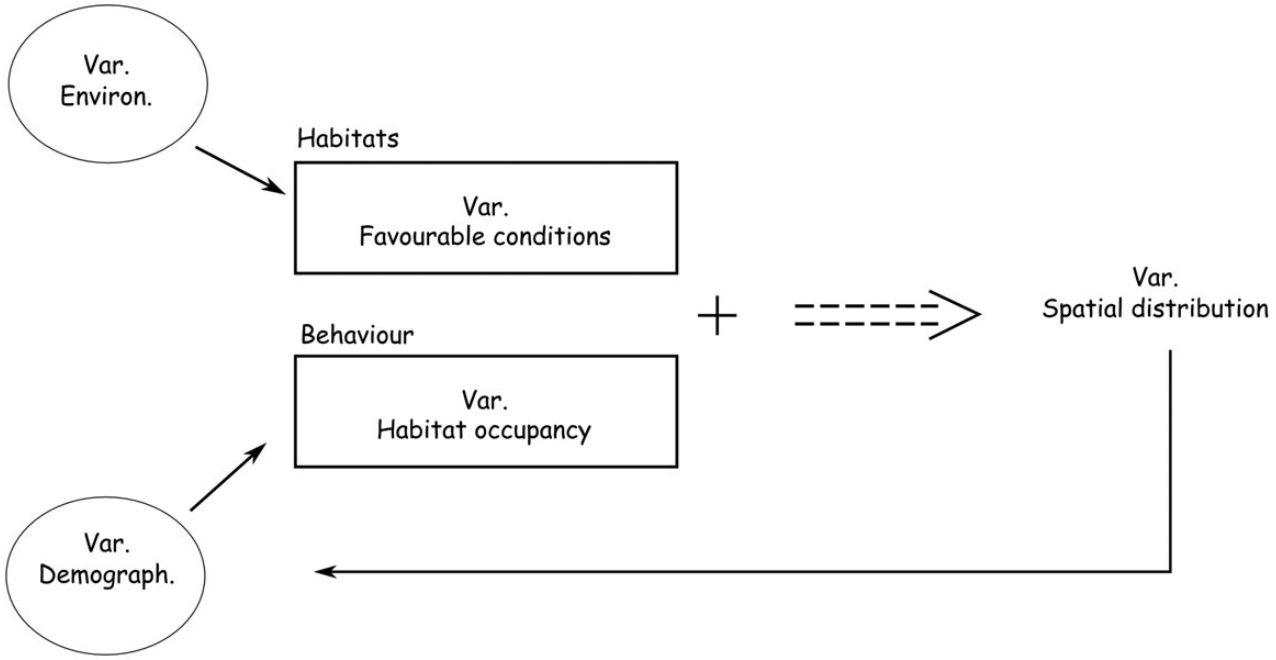

Figure 1. Schematics of the variability in spatial distributions. Environmental and population conditions affect habitats and behaviour, respectively, which influence spatial distribution. In return, the spatial organization of the population also affects its dynamics and thus feeds back into the loop of drivers affecting population behaviour.

a feedback (Figure 1), implying that a particular spatial spawning configuration is associated with a given recruitment regime. Such hypothesis is perhaps more relevant to short-lived species, which show greater variability in their spatial and demographic dynamics. Indeed, sensitivity analyses using coupled physical–biological models demonstrate the importance of initial spawning conditions on larval dispersion and survival for European anchovy (Huret *et al.*, 2010; Ospina-Alavrez *et al.*, 2013). Here the hypothesis is tested by identifying particular types of spawning spatial distributions in fisheries survey dataseries. We suggest a space–time method for doing so, that extracts principle spatial modes in the distribution. Further, we showcase how the method offers the possibility to relate these principal spatial modes to covariates obtained at different spatial resolutions.

To characterize and understand the variability of fish spatial distributions over time, this study intends to showcase the usefulness of applying the space–time decomposition method of empirical orthogonal functions (EOF: Preisendorfer, 1988) on fisheries survey dataseries. In contrast to habitat models where focus is on explaining the mean distribution and its potential change with external drivers only, we here focus with EOFs on characterizing the observed variability around the mean and explaining it with both external and internal population drivers.

In a fisheries management context, characterizing the variability in spatial distributions and understanding their consequences is important for at least two reasons. Marine protected areas (MPA) are often designed based on habitats of particular life history stages (Le Pape *et al.*, 2014). However, variation in the spatial occupancy across years may generate uncertainty in the temporal effectiveness of an MPA as the fish may colonize other areas than the MPA (van Keeken *et al.*, 2007). Further, indicators of spatial distributions have been shown to relate to population parameters (e.g. recruitment, demography, mortality) over a large range of stocks (Woillez *et al.*, 2006). Thus, the characterization of how spatial distributions vary over time will add robustness to population diagnostics as well as spatial management plans.

EOFs have long been used in meteorology and physical oceanography to decompose the time and space variability of a time-series of maps. Fisheries survey series now span more than 10 years and thus offer sufficient space–time information on the spatial distributions of fish populations to consider the use of EOFs for analysing the variability in their spatial distributions. Here, we apply this approach to the time-series of European anchovy (*Engraulis encrasicolus*) spatial distributions at spawning time to extract the main features of variability in spatial occupancy. On that basis, we identify major spatial configurations of the spawning population. We then relate these to year-class strength in the subsequent year. We also explain the spawning configurations with population and environmental parameters. In doing so, we showcase how EOFs provide a methodological framework to understand the ecology of population dynamics in its spatio-temporal dimensions.

Material and methods

Method of empirical of orthogonal functions

The method of EOF (Preisendorfer, 1988) is a particular principal component analysis (PCA) applied to a series of gridded maps, which allows to decompose the space–time (residual) variability in the time-series of maps into principal spatial modes and their amplitudes. The decomposition is a linear factorization of spatial components (eigenvectors) that are constant in time and amplitudes (principal components) that are variable in time. The variability around the mean map is thus modelled as the sum of

time-invariant spatial components that are weighted by their time-varying amplitudes:

$$Z(t, s) = \bar{Z}(., s) + \sum_{m=1}^{q} U_m(t) {E_m}^T(s), \quad (1)$$

where:

$Z(t,s)$ is the variable under study at time t and spatial coordinate s, $\bar{Z}(., s)$ the time average at each coordinate s, $E_m(s)$ the eigenvectors or EOFs (principal spatial modes) scaled to unity, $U_m(t)$ the EOF amplitudes (principal components) scaled to $\sqrt{\lambda_m}$, where the λ_m are the q non-null eigenvalues associated with the EOFs.

To achieve the decomposition, the method proceeds as follows. $Z(t,s)$ is a matrix containing the gridded maps as line vectors with similar spatial order, thus having $t = 1, \ldots, N$ lines and $s = 1, \ldots, K$ columns. Each grid cell must be valued. For missing information, interpolation is needed or use of a coarser grid for data averaging. In each grid cell, the cell time average is subtracted, which results in a matrix of anomalies on which to perform the EOF decomposition: $X(t, s) = Z(t, s) - \bar{Z}(., s)$. Matrix $S = X^T X/N$ is then the covariance in space over time and matrix $Sa = XX^T/K$ the covariance in time over space. A principal components analysis of matrix S (or equivalently Sa) leads to computing the eigenvalues λ_m, eigenvectors $E_m(s)$, and principal components $U_m(t)$.

To retain the most meaningful EOFs and interpret their spatial patterns, we used the eigenvalues (overall variance accounted for by the components) and in addition the "local" explained variance (Schrum *et al.*, 2006; Woillez *et al.*, 2010). The "local" explained variance at location s associated with EOF of order m, $\eta_m(s)$, is the proportion of variance across time that $U_m(t)$ and $E_m(s)$ explain at that location:

$$\eta_m(s) = \frac{\mathrm{Var}[Y_m](s)}{\sum_m \mathrm{Var}[Y_m](s)},$$

where $Y_m = U_m(t)E_m^T(s)$.

When the map of local variance shows subregional patterns, the EOF explains variability in these areas and a biological interpretation can be looked for. Further, when the patterns in the EOF (higher/lower values) can be superimposed on that in the local explained variance, the EOF is dynamically relevant in time and the EOF decomposition well suited to capture the space–time variability in the series of maps. Note that the sign of $Y_m(t,s)$ depends on the combination of the EOF and its amplitude: it is positive when $E_m(s)$ and $U_m(t)$ are of the same sign and negative when they are of opposite signs.

Fish survey data

The survey series considered was the yearly spring acoustic survey series PELGAS, 2000–2012, undertaken by IFREMER on board RV "Thalassa" over the French shelf of the Bay of Biscay in May. The survey design is made of parallel transects, orientated perpendicular to the isobaths and regularly spaced 12 nautical miles (n.m.) apart, from 43.5°N to 48.8°N and from coast (10 m depth) to the shelf break. Along the transects, 38 kHz acoustic records are collected continuously by day, at 10 knots (Doray *et al.*, 2010). Opportunistic pelagic trawl hauls are performed depending on the echotraces and provide information on species proportions, length distributions, weight, and age. During night-time, conductivity–temperature–depth (CTD) profiles are performed on a regular grid of stations (Figure 2), providing measurements of environmental condition.

The anchovy population is surveyed in May at its peak spawning time and its distribution is contained inside the surveyed area (ICES, 2010, chapter 8). European anchovy is mature at age 1 at its first spring and spawning starts in all length groups when surface temperature is above 13°C, corresponding in the Bay of Biscay to the onset of seasonal thermal stratification (Motos, 1996). Thus, the surveyed population corresponds to the spawning adults.

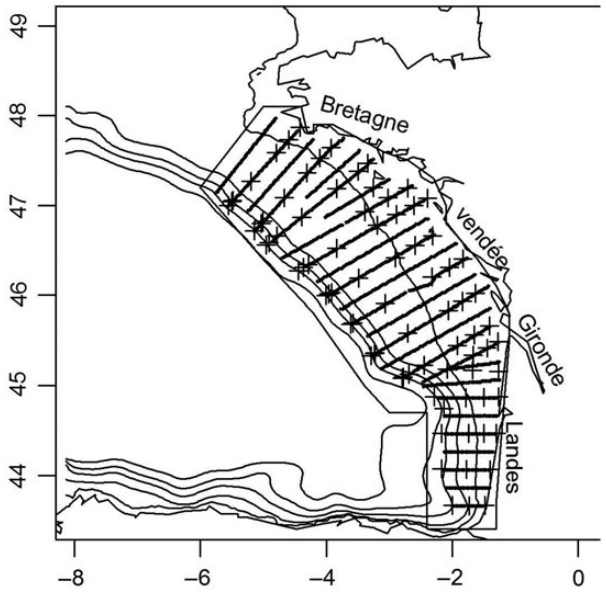

Figure 2. Localization of acoustic transects, CTD stations, and polygon of study.

In the Bay of Biscay, because of the multispecies context and variable schooling characteristics (Massé *et al.*, 1996), the school echotraces cannot be identified to species from their acoustic properties alone with the echosounder currently in use for the assessment. Echotraces are identified to species at a coarser spatial resolution based on pelagic trawl hauls, which are performed over several n.m. to capture aggregations of echotraces. The combination of the trawl haul data with the acoustic data allow to convert acoustic backscatter into fish abundance by species (Simmonds and MacLennan, 2005, section 9; Doray *et al.*, 2010). The resulting data are abundance (tonnes) by species n.m.$^{-2}$ for each n.m. along the survey track. The survey series contains rare but very high values, which may mask the regional patterns of variability. There is uncertainty in these very high values because of uncertainty in allocating acoustic backscatter to species (measurement error) and because extreme biological aggregation is rare and thus little predictable. Thus, we followed a practice used in geoscience in the study of ore deposits when measured concentrations exceed the capability of the measurement device (Rivoirard *et al.*, 2012). Values over a threshold, q, are truncated to that threshold [if $Z(x,t) > q$, then $Z(x,t) = q$]. The threshold considered here was 200 t n.m.$^{-2}$, which corresponded to reported maximum concentration of anchovy school aggregations (2 t per school, 5 schools km^{-1}: Massé *et al.*, 1996; Petitgas *et al.*, 2001). Values >200 were rare with a frequency of 0.002 in the dataset. We considered the spatio-temporal series of these (truncated) data, with one survey per year, 2000–2012.

Block averaging

Before the EOF analysis, the data were averaged by block over a grid, which was the same in all years. The grid mesh size selected was 0.4° in latitude and 0.4° in longitude, with origin x_0 at 43°N and 6°W.

The mean in block (i, j, x_0) was the simple average of the data inside the block and it was positioned at the location of the block centre. The limits of block (i, j, x_0) depend on the origin x_0 and therefore, the samples involved in computing the block mean. To decondition the block means from the grid origin, the point origin x_0 was randomized in the lowest left corner block one 100 times as in Petitgas *et al.* (2009). At each randomization k, the grid origin x_k varies and the mean in block of rank (i, j) is computed. Each block of rank (i, j) has then one 100 means associated with it. The mean of all 100 means was then calculated and positioned at the centre of the block (i, j, x_0). This blocking procedure results in an implicit kernel-like interpolation where each data take part in the block average with a frequency (weight) depending on its distance to the block centre. Finally, blocks which were not valued due to lack of samples in a given year were omitted in the analysis as well as the blocks which had their centre point outside the polygon defining the survey area (Figure 2). Several trials with different mesh sizes were undertaken and the mesh retained was the smallest one for which the first two principal components explained more than 50% of total variance. The grid mesh size retained allowed a reasonable compromise between enough fine scale smoothing and sufficient large and mesoscale details. Finally, instead of biomass per cell, we used the percentage biomass: $P(t, s) = 1000Z(t, s)/\sum_s Z(t, s)$. Matrix $P(t,s)$ was then centred by column to form matrix X.

Typology of maps

The EOF decomposition being a principal components analysis where the years are the lines (individual observations) and the map grid cells the columns (variables), it is possible to perform a hierarchical clustering of the years, applying a clustering procedure on the matrix of distances between years in the factorial space of the first p components ($p < K$). The clustering was performed using the Ward criterium (minimize the intragroup variance), which aims at finding compact groups. The average map in group g, $\bar{Z}_g(., s)$, was estimated by taking the simple average of maps Z for the years of group g. The clustering resulted in defining types of spatial distributions, thus summarizing the meaningful variability in the series of maps.

Explicative covariates

In a classical habitat statistical approach (e.g. Le Pape *et al.*, 2014), the fish distribution map (response) and the maps of (explanatory) covariates are produced on the same spatial grid and each point in the maps contribute to an overall correlation model linking fish presence to covariates. Here, the EOF framework proposes another approach. The EOF patterns of variability are defined empirically from the data and are invariant in time. The variation over time in the spatial distribution is taken in charge by the variations in the weights of the EOF patterns as given by the EOF amplitude time-series. We thus focused on explaining the EOF amplitudes time-series with covariates. Based on our conceptual approach (Figure 1), covariates considered were population parameters affecting population behaviour via density-dependent processes and hydrological conditions affecting habitat suitability.

Population parameters

Following Cotter *et al.* (2009), we considered three indices to characterize population status: total (spawning) biomass and two length distribution percentiles. Population biomass was estimated from the survey as reported to ICES (2012). The length distribution percentiles were the 25 and 75% percentiles. The length distribution in the population was estimated as follows. The (local) frequency distribution of length in a trawl haul was weighted by fish biomass along the acoustic transects within a radius of 10 n.m. from the traul haul position. The population length distribution was the weighted average of all local distributions.

Hydrological indices

Hydrological condition was characterized using indices calculated from the CTD profiles at the stations (Figure 2). We considered four indices (Table 1) to characterize typical features on the shelf in spring (Huret *et al.*, 2013): surface and bottom temperature, a water column stratification index, and a river plume index. These were calculated following the procedures detailed in Huret *et al.* (2013). Surface values were taken at 7 m depth. Bottom values were 5 m above bottom. The deficit of potential energy, Dep, is the energy required to homogenize the water column. The greater the Dep value, the greater the stratification. It was calculated from surface to 60 m depth for the profiles where bottom depth exceeded 60 m. The equivalent freshwater height, H_{fw}, measures the height of accumulated freshwater considering a reference seawater salinity (35.85 psu: mean bottom salinity). Compared with surface salinity, H_{fw} is less affected by vertical mixing and thus reflects better the past history of river discharge over a few months. The indices were calculated at each station and the overall spatial mean was calculated to obtain time-series of mean indices.

Correlation between amplitudes of EOFs and covariates

The time-series of each amplitude $U_m(t)$ associated with EOF of order m was linearly regressed on that of the different covariates $X_j(t)$: $U_m(t) = \sum_{j=0}^{p} a_j X_j(t)$; where $X_0(t) = 1$. For each m, the most probable model was selected using Akaike's information criterion (AIC: Burnham and Anderson, 2002). For each m, we considered the 127 possible models, given the set of the seven covariates. For each m, the model retained was that with the lowest AIC value. To identify the most explanatory covariates (i.e. those involved in the most probable models), we computed their relative importance weights. For that, each covariate j involved in model i explaining amplitude U_m was attributed the model probability $p_m(i)$ as deduced from its AIC value or zero if model i did

Table 1. Hydrological indices calculated using CTD profiles collected during the PelGas cruises.

Index	Units	Formula
Stratification index		
Deficit of potential energy (Dep)	kg m^{-1} s^{-2}	$\text{Dep} = \frac{1}{H}\int_{-H}^{0}(\bar{\rho} - \rho(z))gz\,dz;\ \bar{\rho} = \frac{1}{H}\int_{-H}^{0}\rho(z)\,dz$
River plume index		
Equivalent freshwater height (H_{fw})	m	$H_{fw} = \int_{-H}^{0}\frac{s_0 - s(z)}{s_0}dz;\ s_0 = 35.85$
Hydrological indices		
Surface (7 m) temperature (T_s)	°C	
Bottom temperature (T_b)	°C	

ρ, water density; z, depth; H, bottom depth or 60 m when bottom depth was greater; g, gravitational acceleration; s_0, reference seawater salinity.

not involve the covariate. The relative importance weight of covariate j in explaining amplitude U_m was the sum over all model probabilities: $\sum_{i=1}^{127} p_m(i)$, where $p_m(i) = \exp[-0.5\Delta(i, m)]/\sum_{i=1}^{127} \exp[-0.5\Delta(i, m)]$, with $\Delta(i,m) = \mathrm{AIC}(i,m) - \min(\mathrm{AIC}(i,m))$. The relative importance weight varies between 0 and 1. Typically, strong explanatory covariates will have a relative importance weight around 0.9, moderately explanatory between 0.6 and 0.9, and weakly explanatory between 0.5 and 0.6. Below 0.5, covariates will often be little relevant.

Results

Patterns of variability

The average distribution over all years (Figure 3) shows anchovy to be mainly located south of 46°N with low abundance in the north, mainly in coastal areas. The major concentration is located off the Gironde estuary (45–46°N, 1.5–2°W), along the coast and on the shelf off Landes. The principal spatial modes of yearly variability around the mean were extracted using the EOF decomposition technique. Four principal components were retained ($m = 4$). Each explained more than 10% of total variance (Table 2) and showed a strong spatial pattern in their EOF as well as in the map of local explained variance (Figures 4 and 5). The first EOF (Figure 4) captured the variability along the coast of Landes and off Gironde estuary: the lower the abundance along the coast of Landes, the higher the concentration off Gironde. The second EOF (Figure 4) captured a large-scale pattern where less abundance in the south was associated with more abundance in the north. More specifically, EOF2 captured the variability at the shelf break off Landes, the central part of the shelf off Gironde and Bretagne: the lower the abundance at the coast off Gironde and along the shelf break off Landes, the greater it was in Bretagne and in the central part of the shelf off Gironde. The third EOF (Figure 5) captured the variability along the coast of Vendée and also in the area where the shelf break is curved (45°N, 2°W): the lower the abundance on the shelf off Landes, the greater it was along the coast of Vendée. The fourth EOF (Figure 5) captured the variability along the shelf break north of 45°N: the lower the abundance at the coast off Vendée and at 45°N, the higher at the shelf break.

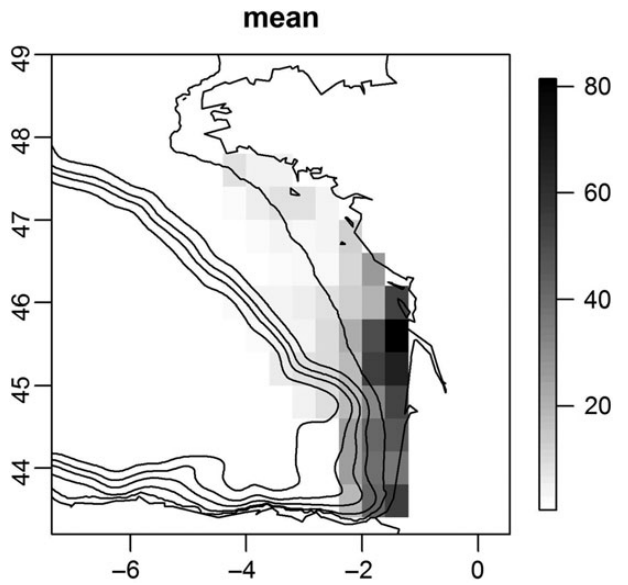

Figure 3. Mean distribution map, 2000–2012. The map represents the average percentage (× 1000) of population biomass in grid cells of 0.4° × 0.4°.

Typology of maps and their relationship with recruitment

Using the first four principal components associated with the EOFs described previously, three groups of maps were identified by hierarchical clustering (Figure 6). The differences among the average maps of each group denoted strong differences in the spatial distributions over the years (Figure 7). Group 1 (Figure 7: maps in years 2000, 2001, 2008, 2009, 2011, 2012) corresponded to the largest spatial extension. The distribution extended in the northern part, on the shelf break north of 45°N and along the coast of Vendée and Bretagne, although the largest concentration was off Gironde and on the shelf of Landes. In constrast, Group 2 (Figure 7: maps in years 2002, 2005, 2010) corresponded to the smallest spatial extension and showed higher concentrations in two areas, off Gironde and at the shelf break off Landes. Group 3 (Figure 7: maps in years 2003, 2004, 2006, 2007) corresponded to a coastal spatial distribution with high concentrations along the coast south of 46°N. This typology of spatial patterns at spawning time related to biomass in the current year and to recruitment (age 1 fish) in the subsequent year (Figure 8): very coastal spawning distributions of type G3 occurred when biomass was low only and at low biomass level, spawning distributions G3 and G2, which had smaller spatial extension than G1, were never followed by high recruitment.

Correlates of amplitudes of EOFs

Based on the AIC criteria, the selected models showed high R^2 ranging 0.48–0.70 (Table 3) and made sense biologically. Temperature alone explained amplitude of EOF 4, which characterized variability along the shelf break. In contrast, amplitude of EOF 3 (which characterized variability along the coast off Vendée) was explained by population parameters only. EOF amplitudes 1 and 2 were explained by a combination of population and environmental indices. Population biomass intervened as covariate of the amplitudes of the three first EOFs, meaning that variability in the distributions was strongly density-dependent. Also, the length distribution influenced the amplitudes of EOFs 1 and 3, which both involved variability in coastal waters. Bottom temperature and water column stratification were covariates of the amplitudes of EOFs 1 and 2, which both involved expansion to the north and variability on mid-shelf. It is noteworthy that the river plume index was never selected as a covariate. The relative importance weights of the covariates (Table 4) confirmed that the covariates in the models selected were the most explanatory ones. Yet, for the amplitude of EOF1, the most probable model (Table 3) involved the stratification index (Dep) which had a low importance weight (Table 4). The model without the stratification index (i.e. with SSB, q75, and T_b

Table 2. Cumulated per cent total variance explained by the principal components of the EOF decomposition.

PC	1	2	3	4	5	6	7	8	9	10	11	12	13
% var	0.30	0.50	0.65	0.77	0.84	0.89	0.93	0.95	0.97	0.98	0.99	1.00	1.00

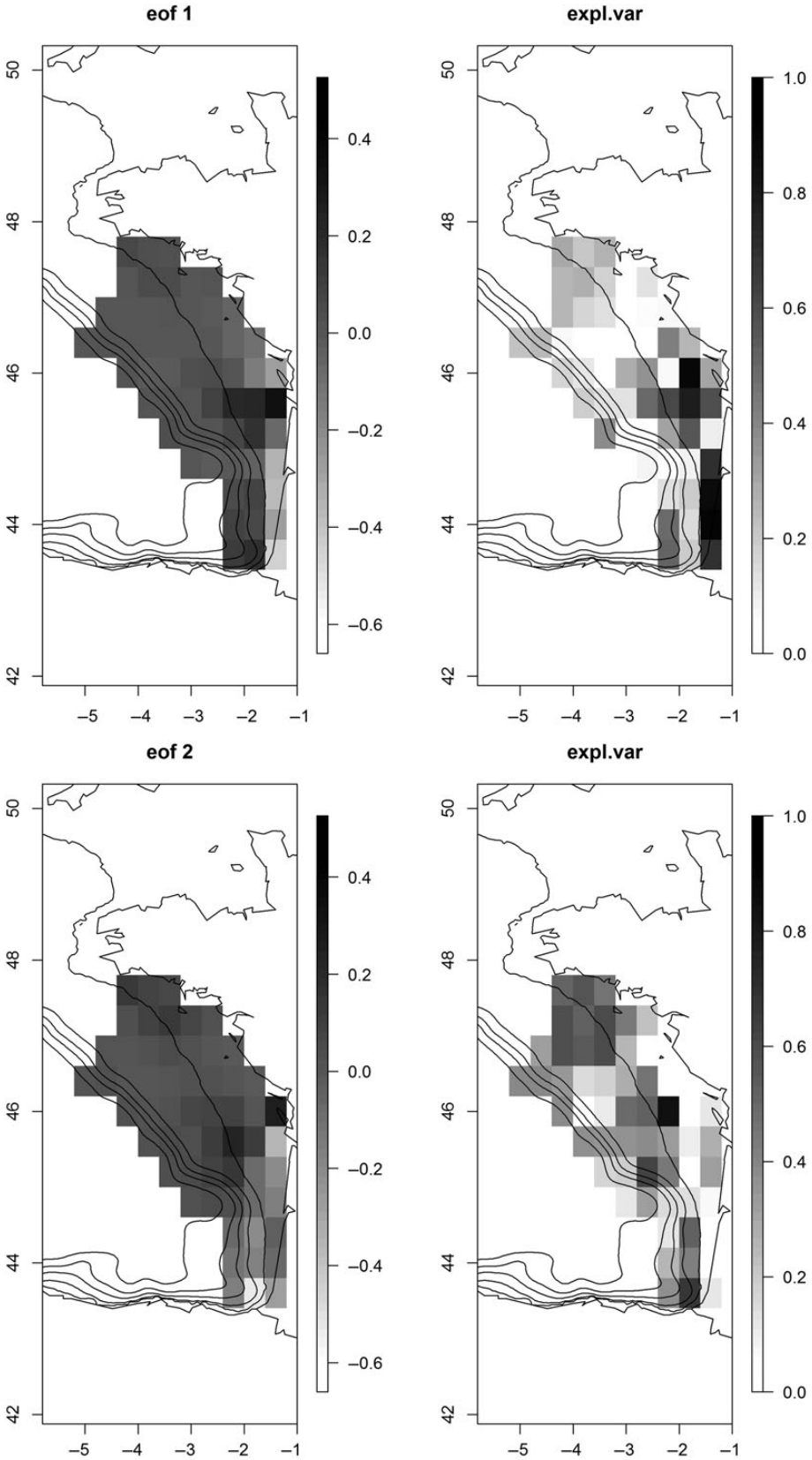

Figure 4. EOFs 1 and 2 (left) and their associated local explained variance (right). EOF1 (top) corresponds to the following pattern of variability: when there is less fish at the coast off Landes (negative values), there is more fish off Gironde (positive values) and *vice versa*. For EOF2 (bottom), the pattern of variability is the following: when there is less fish in southern Biscay (negative values), there is more fish in northern Biscay (positive values) and *vice versa*.

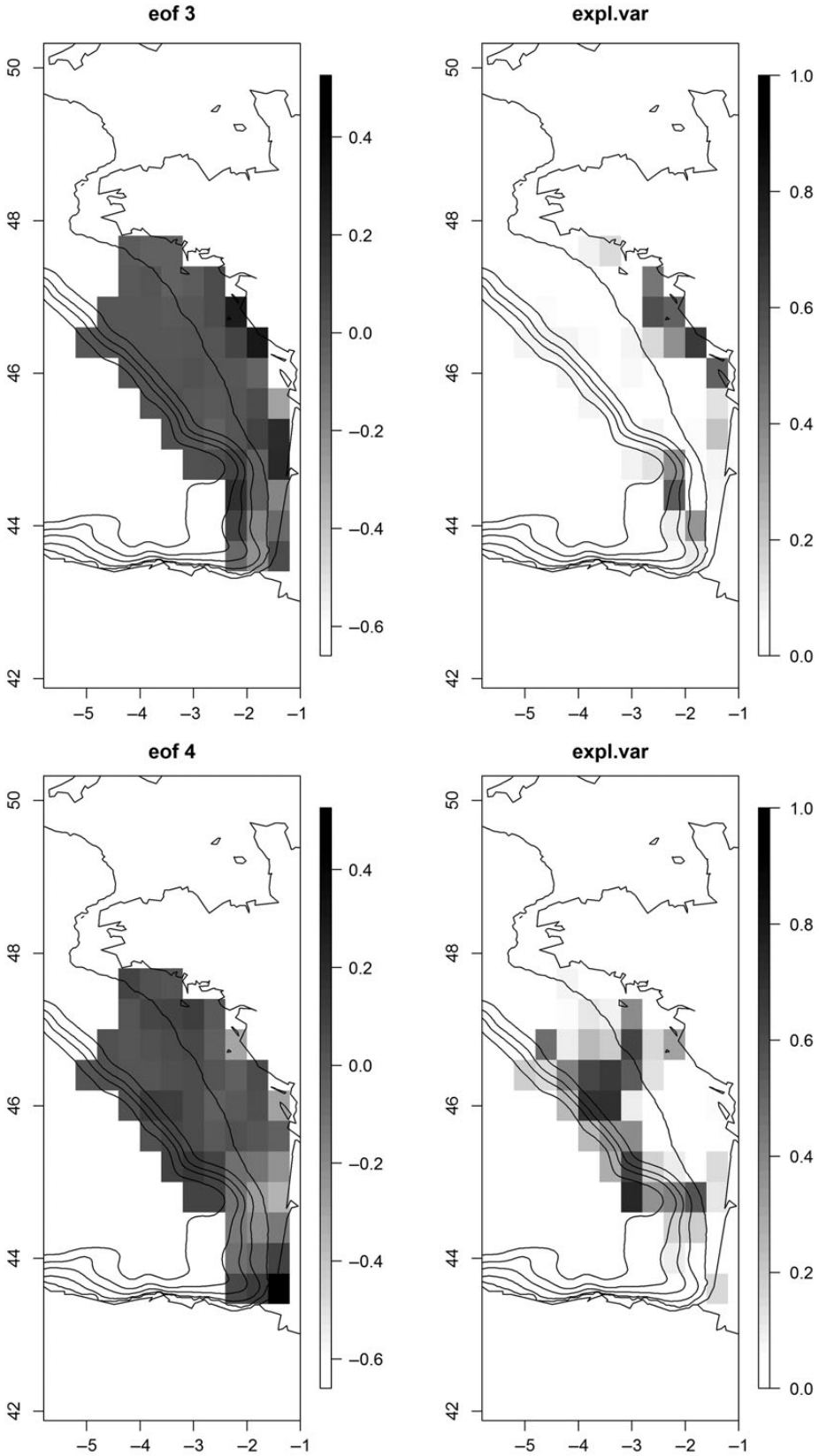

Figure 5. EOFs 3 and 4 (left) and their associated local explained variance (right). EOF3 (top) corresponds to the following pattern of variability: when there is less fish on shelf off Landes (negative values), there is more fish at coast off Vendée (positive values) and *vice versa*. For EOF4 (bottom), the pattern of variability is the following: when there is less fish on shelf off Landes (negative values), there is more fish at the shelf edge in North (positive values) and *vice versa*.

only) was ranked second most probable with a probability of 0.85 and thus was nearly as probable as the retained model (with SSB, q75, T_b, and Dep: Table 3). The stratification index played a slightly minor role relative to the other three covariates in explaining the amplitude of EOF1. For the amplitude of EOF3, the most probable (Table 3) model involved SSB which also had a low importance weight (Table 4). The model without SSB (with q25 and q75 only: Table 3) was ranked second best model but with the low probability of 0.49, meaning that SSB should nevertheless be retained as a covariate. In all, biomass (SSB), proportion of large fish (third quartile q75), bottom temperature (T_b), and water column stratification (Dep) were the most explanatory covariates of the changes over time in the spatial distributions.

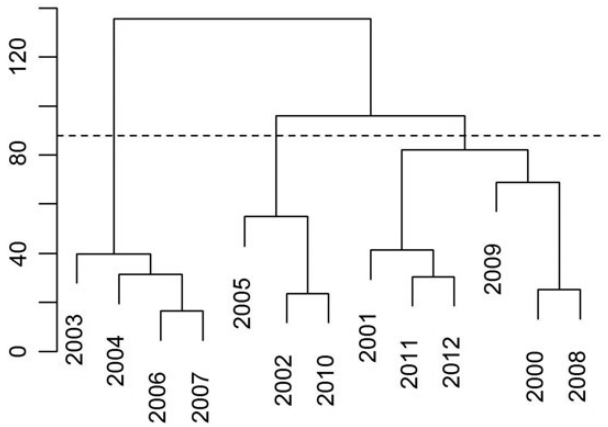

Figure 6. Typology of maps (hierarchical clustering) based on the first four principal components associated with the EOFs. Three groups were retained.

Discussion

The EOF decomposition characterized the variability in the spatial distributions over time by extracting time-invariant principal spatial modes and their time-varying amplitudes. The EOF decomposition served to classify maps. The typology of the spawning spatial distributions related to population subsequent recruitment. The time-series of amplitudes were explained by population parameters (abundance, length distribution) and environmental conditions (bottom temperature, water column stratification). Thus, changes in the spatial distribution were modelled depending on how the time-varying amplitudes (weights of the principal EOF patterns) varied with covariates.

How meaningful are EOFs

The EOFs characterize spatial patterns of variability around the mean map that are constant in time. It is their amplitudes that vary over time not the EOFs. Such mathematical decomposition is not always suited to characterize complex natural space–time variability and we here discuss how to acknowledge their suitability to the case study. The EOF decomposition is little adapted to situations where areas of variability change location over time. The local explained variance (Schrum *et al.*, 2006) is a way to test for that. If the local explained variance displays similar spatial patterns than the EOFs, the EOFs are then dynamically relevant in time. Here, EOFs and their local explained variance showed similar maps. Another point of consideration is that the mathematical property

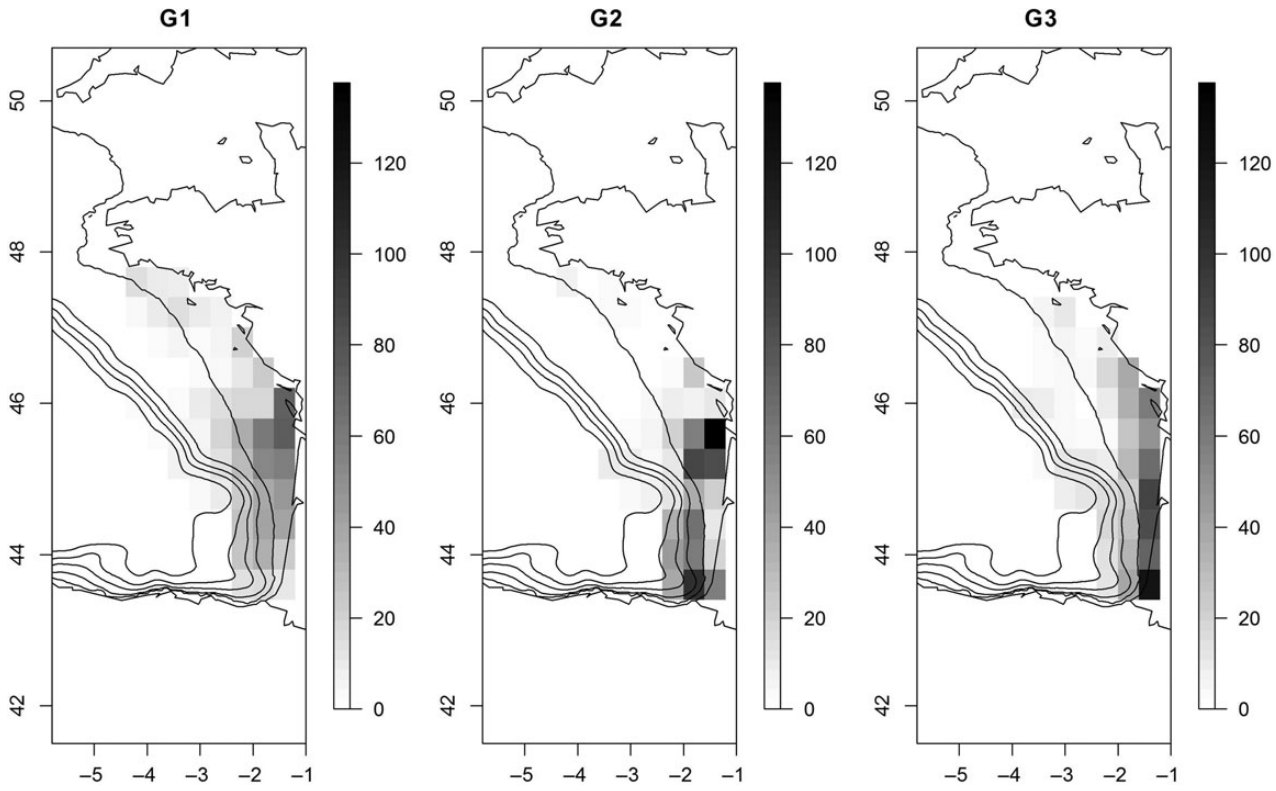

Figure 7. Mean map in each of the three identified groups of maps: G1 (years 2000, 2001, 2008, 2009, 2011, 2012), G2 (years 2002, 2005, 2010), and G3 (years 2003, 2004, 2006, 2007).

of spatial orthogonality in the EOF decomposition may generate artifactual patterns in the spatial modes with little physical interpretation, such as, for example, dipole structures (Dommenget and Latif, 2002). It is therefore key to be able to critically interpret the spatial patterns in the EOFs by relating them to physical or biological phenomena. To increase interpretation, orthogonal (Varimax) rotation of EOFs has been proposed to extract more locally defined spatial patterns (e.g. Richman, 1986). But this approach has its limitations: the ability to extract global spatial patterns in the modes is lower (Dommenget and Latif, 2002) and the rotation may introduce correlation between amplitudes of rotated modes (Mestas-Nuñez, 2000) and thus reduce the ability to reconstruct the data and classify maps. Here, the EOFs identified local areas that were meaningful for the stock (coast, Gironde, shelf break), making unnecessary to find more localized patterns by rotation. The EOF decomposition was thus considered suitable here to characterize the space–time variability in the series of anchovy spawning spatial distributions.

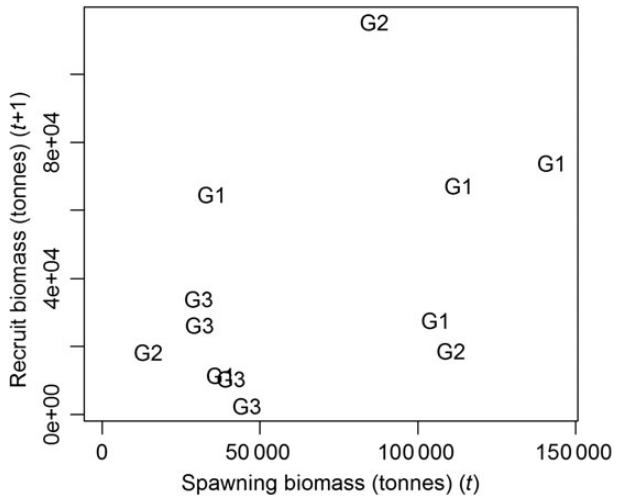

Figure 8. Relationship between recruitment in year ($t + 1$), spawning-stock biomass, and spawning distribution map type (G1–G3) in year (t). At low spawning biomass level, maps of type G3 and G2 in year (t) are never followed with a high recruitment in year ($t + 1$). Map of type G3 is observed only at low biomass level.

Data smoothing on a grid before the EOF decomposition

EOFs are best interpretable when their spatial patterns develop over several grid cells at regional or subregional scale. The grid mesh size over which to block average the data should thus be adequate. Here, we used the amount of variance in the EOFs to define the grid mesh size. The scale in the EOF patterns was greater than 1°latitude × 1°longitude, which was compatible with the mesoscale/subregional aggregation pattern of schools in the area (Petitgas, 2003). The dataset contained rare extreme values, which were uncertain given current knowledge of schooling aggregation in the area (Petitgas *et al.*, 2001). To deal with them, the data were truncated. The truncation threshold was not defined statistically but based on known schooling behaviour, similarly as for filtering unreliable information. Morfin *et al.* (2012) took a different approach to study the variability of spatial patterns: (i) they worked on log-transformed data to diminish the influence of high values; (ii) they interpolated the transformed data by kriging on a grid of small mesh size, which amounts to smoothing. In contrast, our analysis provides guidelines to work on the raw data and makes full use of the local explained variance to interpret the EOFs and their dynamics in time.

Covariates

The regression of EOF amplitudes on series of explicative covariates was a flexible approach as it allowed to relate spatial patterns at regional and subregional scales (EOFs) in the fish distribution to covariates obtained at other spatial resolutions (e.g. population scale). The three major patterns of variability in the spatial distribution (Table 3) amounted to range expansion to northern Biscay, coastal distribution, aggregation off Gironde estuary. Their relative importance was controlled by a combination of environmental conditions and population parameters. Population biomass, population length distribution, bottom temperature, and water column stratification were influential in making the spatial distributions vary, which agrees with the literature. Range expansion with population abundance has been observed in many fish stocks (Shepherd and Litvak, 2004) and was originally observed on anchovy in the California current (MacCall, 1990). Also, anchovy in the Bay of Biscay shows a gradient in length from coast to offshore (Petitgas *et al.*, 2003), the smaller fish being more coastal. It is therefore not surprising that the anchovy distribution varies with population abundance and length distribution and our analysis models how this happens. Also bottom temperature and water column

Table 3. Selected linear regression models of EOF amplitudes U_m on covariates.

EOF	EOF1	EOF2	EOF3	EOF4
EOF description	Less at coast off Landes, more off Gironde	Less in southern Biscay, more in North	Less on shelf off Landes, more at coast off Vendée	Less on shelf off Landes, more at shelf edge in North
Amplitudes	U_1	U_2	U_3	U_4
Intercept	97.44	−201.3	196.3	−127.432
SSB	0.0005875	0.0002659	0.0002982	–
q25	–	–	−24.87	–
q75	12.54	–	31.93	–
T_s	–	–	–	8.789
T_b	−29.93	12.98	–	–
Dep	0.4274	0.5136	–	–
H_{fw}	–	–	–	–
R^2	0.68	0.70	0.43	0.52

Covariates are: SSB, stock biomass as estimated by the survey (t); q25 and q75, the 25 and 75 percentiles of the length distribution (cm); T_s and T_b, the surface and bottom temperature (°C); Dep, the index of stratification; H_{fw}, the index of river plume (Table 1). Values tabled are the coefficients estimated for each covariate. R^2 is the variance explained by the linear model (1 − residual variance/total variance).

Table 4. Relative importance weights of covariates over the 127 candidate linear models for each EOF amplitude U_m.

	U_1	U_2	U_3	U_4
SSB	0.97	0.80	0.47	0.32
q25	0.47	0.40	0.63	0.47
q75	0.55	0.45	0.53	0.40
T_s	0.40	0.44	0.32	0.81
T_b	0.87	0.60	0.36	0.36
Dep	0.43	0.62	0.30	0.42
H_{fw}	0.38	0.38	0.34	0.48

Covariates are defined in Table 3.

stratification were influential. During daytime, anchovy forms schools well below the thermocline generally 10–20 m above the bottom (Massé, 1996), which may explain why bottom temperature may influence habitat suitability and thus anchovy spatial distribution. Water column stratification is under the influence of river plumes and warming surface temperature. River plumes are related to early season plankton production (Labry *et al.*, 2001) and warming surface temperature triggers spawning (Motos, 1996) that occurs at night close to surface. Water column stratification is thus also naturally involved in determining spawning habitat suitability.

EOFs and typology of maps

Based on the EOF decomposition, we classified the varying spawning distributions into three major types of maps: distributions extending over many habitats including northern parts of Biscay (type G1), distributions contracted on two core habitats in south Biscay (type G2), and distributions limited to coastal areas in south Biscay (typeG3). At low biomass level, the distributions of type G3 and G2 that were more contracted spatially and coastal were never followed by a high recruitment in the following year. The coastal distribution of type G3 occurred when biomass was lower, fish length smaller, and water column stratification greater. In these circumstances, total annual fecundity can be expected to be lower because of lower spawning biomass and a shorter spawning period as predicted by bioenergetics (Pecquerie *et al.*, 2009). Also when too high, water column stratification can be detrimental for larval survival (Allain *et al.*, 2007). At low biomass level, the spawning distribution of type G2 was associated with larger fish and colder bottom temperature. These conditions could also result in lower total fecundity over the restricted G2 habitats. Further, the spatial initial conditions for larvae drifts will be more coastal and with less spatial extension for types G3 and G2 than for G1 and this may also influence larval trajectories and survival probability. Whether G2 and G3 spawning types at biomass level can be associated with low subsequent recruitment could be mechanistically tested with a coupled biophysical larval model predicting larval transport and survival. In doing so, a given spawning map type could be used as initial condition to the larval model, following the approach of Ospina-Alavrez *et al.* (2013). Using different egg maps as initial condition for their transport larval model as derived from survey data, they found significant differences in larval drift trajectories.

Dynamic update of maps

Variability in the spatial distributions under both population and environmental conditions is a useful knowledge for increasing robustness in spatial management measures. Based on the regressions of EOF amplitudes on covariates, spatial distributions could be predicted to update dynamically the spawning maps depending on combined population and environmental scenarios. Robustness of spatial management measures could then be tested on that basis. The study provides a typology of spawning distributions that can be input to dynamic population spatial models.

Acknowledgements

This study was jointly supported by Ifremer, the Reproduce project of the EraNet MariFish and the Seaman project of the EraNet SeasEra. We thank Paul Bourriau for taking care of the CTD casts and the equipments and Patrick Grellier and Erwan Duhamel for the individual fish measurements. We thank the crew of RV Thalassa for operating the vessel during the Pelgas surveys. The survey series is partly supported by the EU Fisheries Data Collection Framework. We thank the editor and two anonymous referees for helping improve the paper.

References

Allain, G., Petitgas, P., and Lazure, P. 2007. The influence of environment and spawning distribution on the survival of anchovy (*Engraulis encrasicolus*) larvae in the Bay of Biscay (NE Atlantic) investigated by biophysical simulations. Fisheries Oceanography, 16: 506–514.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modelling, 200: 1–19.

Burnham, K., and Anderson, D. 2002. Model Selection and Multimodel Inference: a Practical Information-Theoretical Approach. Springer, New York.

Cotter, J., Mesnil, B., Witthames, P., and Parker-Humphreys, M. 2009. Notes on nine biological indicators estimable from trawl surveys with an illustrative assessment for North Sea cod. Aquatic Living Resources, 22: 135–153.

Dommenget, D., and Latif, M. 2002. A cautionary note on the interpretation of EOFs. Journal of Climatology, 15: 216–225.

Doray, M., Massé, J., and Petitgas, P. 2010. Pelagic fish stock assessment by acoustic methods at Ifremer. http://archimer.ifremer.fr/doc/00003/11446/.

Guisan, A., and Zimmermann, N. 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135: 147–186.

Huret, M., Petitgas, P., and Woillez, M. 2010. Dispersal kernels and their drivers captured with a hydrodynamic model and spatial indices: a case study on anchovy (*Engraulis encrasicolus*) early life stages in the Bay of Biscay. Progress in Oceanography, 87: 6–17.

Huret, M., Sourisseau, M., Petitgas, P., Struski, C., Léger, F., and Lazure, P. 2013. A multi-decadal hindcast of a physical–biogeochemical model and derived oceanographic indices in the Bay of Biscay. Journal of Marine Systems, 109–110: S77–S94.

ICES. 2010. Life cycle patterns of small pelagic fish in the north east Atlantic. ICS Cooperative Research Reports, No. 306.

ICES. 2012. Report of the Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA). ICES CM 2012/ACOM: 16.

Labry, C., Herbland, A., Delmas, D., Laborde, P., Lazure, P., Froidefond, J-M., Jégou, A-M., *et al.* 2001. Initiation of winter phytoplankton blooms within the Gironde plume waters in the Bay of Biscay. Marine Ecology Progress Series, 212: 117–130.

Le Pape, O., Delavenne, J., and Vaz, S. 2014. Quantitative mapping of fish habitat: a useful tool to design spatialised management measures and marine protected area with fishery objectives. Ocean and Coastal Management, 87: 8–19.

MacCall, A. 1990. Dynamic Geography of Marine Fish Populations. University of Washington Press, Seattle.

Massé, J. 1996. Acoustic observations in the Bay of Biscay: schooling, vertical distribution, species assemblages and behaviour. Scientia Marina, 60 (Suppl. 2): 227–234.

Massé, J., Koutsikopoulos, C., and Patty, W. 1996. The structure and spatial distribution of pelagic fish schools in mutlispecies clusters: an acoustic study. ICES Journal of Marine Science, 53: 155–160.

Mestas-Nuñez, A. 2000. Orthogonality properties of rotated empirical modes. International Journal of Climatology, 20: 1509–1516.

Morfin, M., Fromentin, J-M., Jadaud, A., and Bez, N. 2012. Spatio-temporal patterns of key exploited marine species in the Northwestern Mediterranean Sea. Plos One, 7: e37907. doi:10.1371/journal.pone.0037907

Motos, L. 1996. Reproductive biology and fecundity of the Bay of Biscay anchovy population (*Engraulis encrasicolus*, L.). Scientia Marina, 60 (Suppl. 2): 195–207.

Ospina-Alavrez, A., Bernal, M., Catalán, I., Roos, D., Bigot, J.-L., and Palomera, I. 2013. Modeling fish egg production and spatial distribution from acoustic data: a step forward into the analysis of recruitment. PLoS One, 8: e73687. DOI: 10.1371/journal.pone.0073687

Pecquerie, L., Petitgas, P., and Kooijman, S. 2009. Modeling fish growth and reproduction in the context of the Dynamic Energy Budget theory to predict environmental impact on anchovy spawning duration. Journal of Sea Research, 62: 93–105.

Petitgas, P. 2003. A method for the identification and characterization of clusters of schools along the transect lines of fisheries-acoustic surveys. ICES Journal of Marine Science, 60: 872–884.

Petitgas, P., Goarant, A., Massé, J., and Bourriau, P. 2009. Combining acoustic and CUFES data for the quality control of fish-stock survey estimates. ICES Journal of Marine Science, 66: 1384–1390.

Petitgas, P., Massé, J., Grellier, P., and Beillois, P. 2003. Variation in the spatial distribution of fish length: a multi-annual geostatistics approach on anchovy in Biscay, 1983–2002. ICES CM 2003/Q: 15.

Petitgas, P., Reid, D., Carrera, P., Iglesias, M., Georgakarakos, S., Liorzou, B., and Massé, J. 2001. On the relation between schools, clusters of schools, and abundance in pelagic fish. ICES Journal of Marine Science, 58: 1150–1160.

Petitgas, P., Secor, D., McQuinn, I., Huse, G., and Lo, N. 2010. Stock collapses and their recovery: mechanisms that establish and maintain life-cycle closure in space and time. ICES Journal of Marine Science, 67: 1841–1848.

Planque, B., Loots, C., Petitgas, P., Linstrøm, U., and Vaz, S. 2011. Understanding what controls the spatial distribution of fish populations using a multi-model approach. Fisheries Oceanography, 20: 1–17.

Preisendorfer, R. 1988. Principal Component Analysis in Meteorology and Oceanography. Elsevier, Amsterdam.

Richman, M. 1986. Rotation of principal components. Journal of Climatology, 6: 293–335.

Rivoirard, J., Demange, C., Freulon, X., Lécureuil, A., and Bellot, N. 2012. A top-cut model for deposits with heavy-tailed grade distribution. Mathematical Geoscience, 44. doi:10.1007/s11004–012-9401-x

Schrum, C., St John, M., and Alekseeva, I. 2006. ECOSMO, a coupled ecosystem model of the North Sea and Baltic Sea: Part II. Spatial–seasonal characteristics in the North Sea as revealed by EOF analysis. Journal of Marine Systems, 61: 100–113.

Shepherd, T., and Litvak, M. 2004. Density-dependent habitat selection and the ideal free distribution in marine fish spatial dynamics: considerations and cautions. Fish and Fisheries, 5: 141–152.

Simmonds, J., and MacLennan, D. 2005. Fisheries Acoustics: Theory and Practice, 2nd ed. Blackwell Science, Oxford.

Van Keeken, O., van Hoppe, M., Grift, R., and Rijnsdorp, A. 2007. Changes in the spatial distribution of North Sea plaice (*Pleuronectes platessa*) and implications for fisheries management. Journal of Sea Research, 57: 187–197.

Woillez, M., Petitgas, P., Huret, M., Struski, C., and Léger, F. 2010. Statistical monitoring of spatial patterns of environmental indices for integrated ecosystem assessment: application to the Bay of Biscay pelagic zone. Progress in Oceanography, 87: 83–93.

Woillez, M., Petitgas, P., Rivoirard, J., Fernandes, P., terHofstede, R., Korsbrekke, K., Orlowski, A., *et al.* 2006. Relationships between population spatial occupation and population dynamics. ICES CM 2006/O: 05.

Handling editor: Valerio Bartolino