Empirical Orthogonal Functions for ecology

Baptiste Alglave, Benjamin Dufée, Said Obakrim, James T Thorson

Empirical Orthogonal Functions for ecology

Baptiste Alglave[1,2], Benjamin Dufée[1,3], Said Obakrim[4], and James T. Thorson[5]

Abstract: Spatio-temporal data are ubiquitous in ecology. As the volume of ecological data increases, they encompass long time periods on large spatial domains, frequently involving multiple species and variables. Developing methods to summarize these data while providing ecological interpretability is crucial for maximizing the insights they offer.

Empirical Orthogonal Functions (EOFs) are a well-documented and widely used method for reducing the dimensionality of spatio-temporal data. First, introduced in the 1950s, EOFs essentially involve performing Principal Component Analysis on spatio-temporal data. Subsequently, a substantial body of literature has developed around EOFs, primarily in the fields of meteorology and climate studies. While a substantial body of research has applied EOFs in meteorology and climate science, their use in ecology is relatively new and remains overlooked, despite their huge potential.

In this paper, we aim at presenting the basics of EOFs and introduce new related methods that address specific ecological questions. To illustrate their potential, we use two different datasets: (1) satellite data for the Normalized Difference Vegetation Index (NDVI) over France, and (2) spatio-temporal predictions from an integrated species distribution model that maps several fish species in the Bay of Biscay.

Since EOFs do not incorporate ecological constraints, their results may not always be directly interpretable from an ecological perspective. In this paper, we develop a method that incorporates a temporal ecological variable into EOFs to improve their interpretability. Using satellite NDVI data, we demonstrate how EOFs results can be informed by including a steady seasonal variable, which helps to better represent a North-South gradient across France.

Finally, EOFs can be a valuable tool for investigating community and ecosystem spatio-temporal dynamics, though this implies to move to a multivariate framework. We illustrate several ways to leverage EOFs for multivariate analysis and highlight shared spatial and temporal patterns among

1 Université Bretagne Sud, IUT Vannes, Department of Data Science, 8 rue Michel de Montaigne, 56000, Vannes, France,
2 Lab-STICC, Team DECIDE, Rue de Saint-Maudé, 56100 Lorient, France,
3 Laboratory of Mathematics in Bretagne Atlantique, Rue de Saint-Maudé, 56100 Lorient, France,
4 Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland, said.obakrim@unil.ch
5 Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA 98112, USA,

several spatio-temporal variables. Using multispecies fish data from the Bay of Biscay, we identify both common temporal trends and shared spatial patterns in reproduction timing and locations. A set of reproducible codes are provided in the manuscript to assist ecologists interested in applying these techniques in practice.

Keywords: spatio-temporal data, dimension reduction, satellite data, NDVI, species distribution model, multivariate analysis, canonical correlation analysis.

1 Introduction

Spatial and temporal data are ubiquitous in ecological research, providing insights across various scales. Survey data, analyzed through multivariate integrated modeling, help understanding species distribution, ecosystem states and community dynamics (Isaac et al., 2020; Zipkin et al., 2021). Citizen science initiatives yield extensive opportunistic datasets over long periods, facilitating studies on phenomena such as bird migration (Martin et al., 2020; Sullivan et al., 2014). GPS data allows to track animal movements at high resolution (Nathan et al., 2022). Remote sensing offers high-resolution information for monitoring land use, habitat suitability, and the impacts of climate change in near real-time (Poggi et al., 2021; Garrigues, Allard, and Baret, 2008). Together, these diverse spatio-temporal data sources are crucial for unraveling spatial ecological processes, their temporal trends, and the interactions affecting ecological processes across various scales.

Typically, integrated species distribution models (SDM) allow to infer the spatio-temporal distribution of species from distinct data sources and provide the common information shared between all these data (Isaac et al., 2020; Seaton, Jarvis, and Henrys, 2024). As a consequence of the massive amount of data that are gathered to feed the model, the spatio-temporal predictions can result in long time series of spatial maps (Alglave, Olmos,

et al., 2024). These still need to be analysed and processed to study the ecological processes structuring these outcomes.

Given the extensive spatial and temporal coverage of these data, dimension reduction (or ordination) methods become critical for synthesizing the information embedded within these spatio-temporal data. Empirical Orthogonal Functions (EOFs) represent the keystone method for performing dimension reduction on spatio-temporal data. They were first introduced by Lorenz (1956), Obukhov (1947), and Fukuoka (1951) and they could be basically summarized as a Principal Component Analysis (PCA) performed on spatio-temporal data where individuals are locations and variables are time steps. Since their first formulation, EOFs methods have entailed a rich literature, particularly in meteorology and climate applications (Hannachi, Jolliffe, and Stephenson, 2007).

Recently, these methods have expanded to ecology. First applications have remained close to their original use by analysing the output of climatic or physical models. For instance, Gedalof, Peterson, and Mantua (2005) explored the relationship between climatic variables and wildland fire data, while other studies analyzed oceanographic and soil variables to understand ecosystem dynamics, marine-terrestrial synchrony, and water conservation for ecological management (Schrum, John, and Alekseeva, 2006; Iida and Saitoh, 2007; Di Lorenzo et al., 2008; Woillez et al., 2010; Le Mezo et al., 2016; Berkelhammer, 2019; Wu et al., 2023).

Researchers have also applied EOFs to species' spatio-temporal distributions (with a specific emphasis on fish and marine ecosystems), linking them to climate variables, particularly in studies of sardines and anchovies in the California Current system (Norton and Mason, 2005) or in the Bay of Biscay (Petitgas, Masse, et al., 2006; Petitgas, 2008; Petitgas, Doray, et al., 2014). Some research used MFA (a derived method of PCA) to explore the spatial structure of the ecosystem based on multi-trophic observations (Woillez et al., 2010; Petitgas, Huret,

et al., 2018; Grandremy et al., 2023). Additionally, Van Audenhaege et al. (2022) examined vent mussel habitats based on imagery data, highlighting small-scale variability influenced by tidal cycles and temperature anomalies.

These methods gained wider recognition through the development of the Vector-Autoregressive Spatio-temporal (VAST) package (Thorson, 2019; Thorson, Anderson, et al., 2024). VAST implements a joint hierarchical model with an underlying EOFs structure. It allows (1) to integrate spatio-temporal data from several ecological variables and species, (2) provide a low rank representation of the ecosystem variability and (3) predict the effect of climate change on the ecosystems (Thorson, Ciannelli, and Litzow, 2020; Thorson, Cheng, et al., 2020; Thorson, Arimitsu, et al., 2021).

At this stage, despite their huge potential these methods remain underused in ecology and mostly unknown in terrestrial ecology. This may be due to their perceived complexity and the lack of clear examples showing how they can be applied to address ecological questions.

In this paper, we provide a comprehensive introduction to the basics of EOFs and how to fairly interpret them. Additionally, to further demonstrate the potential of EOFs for ecology, we introduce two related methods that extend EOFs: (1) a method that informs EOFs with an ancillary ecological variable, and (2) a method that generalizes EOFs to the multivariate case, enabling EOFs analyses on spcies communities or entire ecosystems.

EOFs do not incorporate any ecological constraints, and their results may not always have a direct ecological interpretation due to standard PCA constraints (*i.e.* orthogonality). Using canonical correlation analysis (CCA), we impose an ecological constraint on EOFs to improve their ecological interpretability. Additionally, EOFs have proven to be valuable tools for community modeling and for ecosystem dynamics analysis (Petitgas, Huret, et al., 2018; Thorson, Arimitsu, et al., 2021). Here, we introduce two approaches to perform

multivariate EOFs, corresponding to two types of data representation: a temporally synthetic representation and a spatially synthetic representation. These approaches illustrate the flexibility of the EOFs method for multivariate modeling. While providing different insights into community dynamics, they allow to identify the shared spatial and temporal patterns as well as the distinct habitats or seasons among the species (or ecological variables) included in the analysis.

To illustrate these methods, we apply them to two real ecological case studies. First, we use a terrestrial ecology example based on monthly Normalized Difference Vegetation Index (NDVI) satellite data (MODIS) from 2000 to 2023 (MODIS, 2021). This case study aims at demonstrating the potential of EOFs for terrestrial systems. Second, we use spatio-temporal predictions from an Integrated Spatio-Temporal Species Distribution Model (IST-SDM) (Alglave, Vermard, et al., 2023). EOFs can be a valuable tool to analyse results of an integrated model, providing a synthetic representation of the information gathered from all the datasets. We analyze several demersal species in the Bay of Biscay between 2008 and 2018 at a monthly time step, inferred jointly from scientific survey data and catch declaration data here common sole (*Solea solea*, Linnaeus, 1758), squid (*Loligo vulgaris*, Lamarck, 1798), and whiting (*Merlangius merlangus*, Linnaeus, 1758).

In the following sections, we first summarize the basics of EOFs, introducing (1) the main mathematical concepts underlying EOFs and (2) how results should be interpreted (Section 2). We then present the dataset used to illustrate these methods (Section 3) and apply EOFs methodology on both case studies (Section 4). Next, we introduce a method derived from CCA to inform EOFs with ancillary ecological temporal variables (Section 5), and finally, we demonstrate the flexibility of EOFs for modeling multiple species and performing community analysis (Section 6). Both real case studies are used throughout the manuscript

to illustrate the methods. Additionally, code sets are provided to guide practitioners in performing the analysis in the simplest way possible.

2 Basics of EOFs

EOFs were initially introduced by Fukuoka (1951), Obukhov (1947), and Lorenz (1956) for the analysis and prediction of weather. The method involves extracting the main spatial and temporal patterns that (1) capture the most variance and (2) are orthogonal to each other. In the upcoming subsections, we introduce the notations and outline the principal properties of EOFs through both theoretical explanations and practical applications.

2.1 The raw spatio-temporal field

We denote a space-time process S such that $S(x,t)$ represents the value of the process S at spatial location x and time t. The data under consideration will be a discretized version of this underlying process defined on a regular grid, namely a matrix $\mathbf{S}$ of size $n \times p$, where n denotes the number of discrete spatial positions, and p denotes the number of time steps, with the time stepping denoted by $(t_1, ..., t_p)$.

From these data, for a given spatial location x, the temporal average can be calculated as:

$$\vec{\mathbf{s}}^{t}(x) = \frac{1}{p}\sum_{k=1}^{p} S(x, t_k). \tag{1}$$

The time-centered space-time field, also referred to as the anomaly matrix, is then represented by:

$$\mathbf{S}' = \begin{pmatrix} S'(x_1,t_1) & S'(x_1,t_2) & \cdots & S'(x_1,t_p) \\ S'(x_2,t_1) & S'(x_2,t_2) & \cdots & S'(x_2,t_p) \\ \vdots & \vdots & \ddots & \vdots \\ S'(x_n,t_1) & S'(x_n,t_2) & \cdots & S'(x_n,t_p) \end{pmatrix}$$

where

$$S'(x_l,t_k) = S(x_l,t_k) - \overline{\mathbf{s}}^t(x) \quad \text{for} \quad l = 1,\ldots,n, \quad k = 1,\ldots,p. \tag{2}$$

Examples of both the raw spatio-temporal data and anomalies are available for satellite NDVI data and IST-SDM predictions in Figure 1 and S1. Codes below are provided to compute the anomaly matrix ($\mathbf{S}'$) from spatio-temporal data contained in a matrix $\mathbf{S}$.

Compute the anomaly matrix in R

```
# S is the spatio-temporal matrix (n . p)
# 1/ Compute the temporal average map.
Smean <- sweep( S, MARGIN=2, STATS=colMeans(S), FUN="-" )
# 2/ Center the spatio-temporal dataset and
# obtain the anomaly matrix Sprim
Sprim <- S - outer(rep(1, ncol(S)), Smean)
```

2.2 EOFs formalism: decomposing the spatio-temporal field

2.2.1 The model underlying EOFs

When dealing with EOFs, we assume that the spatio-temporal field can be decomposed into a (purely) spatial term and a (purely) temporal term. This decomposition can be expressed as:

$$S'(x,t) = \sum_{m=1}^{r} u_m(x) \cdot v_m(t) + \epsilon_m(x,t). \tag{3}$$

Here, r represents the number of dimensions of the EOFs with $r \leq \min(n,p)$ being the rank of the matrix $\mathbf{S}'$, $u_m(x)$ denotes the spatial term of the EOFs for dimension m, $v_m(t)$ represents the temporal term of the EOFs for dimension m, and $\epsilon_m(x,t)$ is an error term. In some cases, $\epsilon_m(x,t)$ may not be explicitly defined in the equation (3), and the r dimensions capture all the variability in $\mathbf{S}'$.

The key aspect of EOFs analysis lies in the constraints imposed on the spatial terms $u_m(x)$ or the temporal terms $v_m(t)$ because they are crucial for the inference and interpretation of these indices. The most straightforward and natural constraints involve (1) maximising the variance explained by each spatial pattern $p_m(x)$ which is equivalent to minimizing the global error, defined as $E = \sum_m \sum_x \sum_y \epsilon_m(x,t)$; (2) setting orthogonal constraints for both the spatial and temporal terms.

2.2.2 Writing EOFs in matrix form

To obtain the terms $u_m(x)$ and $v_m(t)$ of equation (3), it is possible to decompose $\mathbf{S}'$ through Singular Value Decomposition (SVD), namely:

$$\mathbf{S}' = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{4}$$

with $\mathbf{U}$ a $n \times r$ matrix and $\mathbf{V}^T$ a $r \times p$ matrix, $r \leq \min(n, p)$ being the rank of the matrix $\mathbf{S}'$. $\mathbf{\Lambda}$ is a $r \times r$ diagonal matrix with non-increasing positive coefficients on the diagonal, denoted as $\mathbf{\Lambda} = \text{Diag}(\lambda_1, ..., \lambda_r)$ with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > 0$. This is essentially the same computation that is performed when conducting PCA on any type of multivariate data.

Here, the columns of $\mathbf{U} = (\mathbf{u}_1, ..., \mathbf{u}_r)$ are referred to as the Empirical Orthogonal Function (EOF) maps of the anomaly matrix $\mathbf{S}'$. In PCA terms, they are often called factors; in the following, they will be referred to as spatial factors.

The columns of $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_r)$ are often referred to as loading factors in PCA terms. Here, they will be called either loadings or temporal loadings of the anomaly matrix $\mathbf{S}'$. These columns contain the terms $v_m(t)$ from Equation (3).

As stated previously, the spatial factors are constructed to be spatially orthogonal, and the loading factors are also temporally orthogonal, meaning that:

$$\begin{aligned} \langle \mathbf{u}_i ; \mathbf{u}_j \rangle_{\mathbb{R}^n} &= 0 \quad i \neq j \\ \langle \mathbf{v}_i ; \mathbf{v}_j \rangle_{\mathbb{R}^p} &= 0 \quad i \neq j \end{aligned} \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the euclidean scalar product.

Finally, $\mathbf{\Lambda}$ contains the singular values of $\mathbf{S}'$ along the diagonal. It represents the amount of

variance captured by each dimension of the EOFs also called the inertia in standard PCA. If we note I_m the variance (or inertia) related to dimension m, we obtain:

$$I_m = \lambda_m^2$$

The matrix decomposition of equation (4) coincides with equation (3) provided the spatial factors $\mathbf{u}_m$ are normalized by the singular values. Below are the codes to perform SVD in R and the corresponding R objects related to the EOFs equations.

Perform SVD on the anomaly matrix

```
E <- svd(Sprim)
SpFactors <- E$u # spatial factors, dimension (n . p)
TempLoadings <- E$v # temporal loadings, dimension (p . p)
SgValues <- E$d # singular values, vector of dimension (p)
```

Finally, note that in standard multivariate statistics, PCA is often combined with clustering analysis (*e.g.*, k-means, Hierarchical Clustering on Principal Components - HCPC) to group individuals into homogeneous clusters. A similar approach can be applied to EOFs to either group locations with similar temporal trends or time steps with similar spatial patterns (Lindegren et al., 2022; Alglave, Olmos, et al., 2024). This will be discussed later, as it can be particularly useful for identifying functional habitats from species spatio-temporal distributions (Alglave, Olmos, et al., 2024).

2.2.3 Interpretation of EOFs

The construction of EOFs discussed in the previous subsection implies that the first EOFs map is the spatial pattern $\mathbf{u}_1$ that captures the most variance in $\mathbf{S}'$. The second EOFs, $\mathbf{u}_2$, captures the next most variance while being orthogonal to $\mathbf{u}_1$.

The loading factors contained in $\mathbf{v}_j$, where $j \in 1, \cdots, r$, are temporal indices. Each is associated with a corresponding spatial factor $\mathbf{u}_j$. When the loading $v_j(t)$ is positive, $\mathbf{S}'$ is distributed according to the spatial factor $\mathbf{u}_j$. Conversely, if $v_j(t)$ is negative, $\mathbf{S}'$ is distributed according to $-\mathbf{u}_j$. It is important to note that the signs of the loading factors and spatial factors are arbitrary, meaning that switching the sign of both $\mathbf{v}_j$ and $\mathbf{u}_j$ does not change the overall model. This label/sign indeterminacy is common in factor models and should be kept in mind when interpreting the results, as the sign of the factors may flip between different analyses or software implementations.

An important consideration is the number of dimensions retained for interpretation. Several rules of thumb can guide this decision. Many are based on the proportion of variance captured by each dimension m of the EOFs denoted $PV_m = \frac{I_m}{\sum_{k=1}^{r} I_k}$ and that is often represented through a scree plot. One approach is to keep the dimensions that capture the variance of more than what a single variable would capture if all the variables were independent. This means retaining dimensions that capture more than 1 out of the number of variables. Another method involves examining the variance graph and selecting dimensions up to the point where there is a noticeable drop in variance explained, indicating a significant change in the slope of the scree plot. Finally, one may choose to retain only the dimensions that are interpretable from an ecological point of view. These criteria are not mutually exclusive. For simplicity, in the following text, we represent only the first two dimensions, with additional dimensions being provided in the Supplementary Material.

EOFs should be interpreted as statistically significant spatial patterns. However, each individual EOFs may not necessarily represent a physically or dynamically relevant spatial pattern (Monahan et al., 2009; Roundy, 2015). For example, a single EOFs dimension may not correspond directly to a specific physical phenomenon; instead, such process might be captured by a linear combination of multiple EOFs. Additionally, there is no constraint separating long-term and short-term signals, so both could be combined within a single dimension. Identifying individual EOFs with underlying processes should be done with great care, always keeping in mind that EOFs capture variance rather than represent an ecological process per se.

An example of EOFs interpretation in climate science is the North Atlantic Oscillation (NAO) index (Hurrell and Deser, 2010), which is often analyzed through EOFs. The NAO is a weather phenomenon in the North Atlantic Ocean characterized by fluctuations in the atmospheric pressure difference at sea level between the Icelandic Low and the Azores High. The NAO has been shown to influence ecological dynamics in both marine and terrestrial systems (Ottersen et al., 2001). The first EOFs dimension of the sea-level pressure field over the North Atlantic typically represents the NAO pattern (Figure 8 in Saeed, Kucharski, and Almazroui, 2023). When the loading factor associated with this dimension is positive, it indicates a strong Azores high and a deep Icelandic low, corresponding to the positive phase of the NAO. Conversely, a negative loading factor indicates the negative phase of the NAO, characterized by a deep Azores low and a strong Icelandic high. This EOFs dimension captures the primary mode of variability in the North Atlantic sea-level pressure field and is crucial for understanding and predicting regional climate variations.

3 Data and case study for illustration

In order to illustrate the various methods, we introduce two spatio-temporal datasets to serve as a case study: (1) NDVI data from MODIS Vegetation Index Products (satellite data on terrestrial ecosystems) and (2) spatio-temporal predictions from a IST-SDM for fishes in the Bay of Biscay (IST-SDM on a marine case study).

3.1 Satellite NDVI data

For NDVI data, we use the one kilometer monthly product available through MODIS (Busetto and Ranghetti, 2016). These are satellite observations data, which provide high-resolution information of the Earth's surface. By capturing reflected light in various spectral bands, these data allows to compute Normalized Difference Vegetation Index (NDVI), a key indicator of vegetation health and density. NDVI values range from -1 to 1, with higher values indicating healthier and denser vegetation. This index is particularly useful for monitoring changes in vegetation over time, assessing drought conditions, and understanding ecosystem dynamics. We filter the data over France and aggregated the data on a grid cell with resolution of 10 km (Figure 1). These data range from 2000 to 2023 with a monthly time step and consist of 288 maps.

NDVI spatio-temporal variability is strongly related to temperature and precipitation, with significant heterogeneity in these relationships across regions (Piao et al., 2003; Vicente-Serrano et al., 2013). Consequently, the seasonal cycle is expected to have a strong influence on NDVI through fluctuations in both precipitation and temperature, with possible heterogeneity depending on local climatic conditions (Eisfelder et al., 2023). Specifically in France, as noted by Eastman et al. (2013), the main pattern should exhibit higher NDVI values during spring, summer, and early autumn, corresponding to higher temperatures

during these periods and increased precipitation in spring and early autumn. In contrast, NDVI is expected to be lower in late autumn and winter due to lower temperatures and vegetation senescence.

3.2 Species distribution model predictions

For the IST-SDM, we use monthly spatio-temporal predictions for three key species in the Bay of Biscay (common sole, *Solea solea* - squids, *Loligo vulgaris* - whiting, *Merlangius merlangus*) spanning a decade from 2008 to 2018 (Figure S1). These predictions were derived from a hierarchical model that integrates data from fishers' declarations and scientific surveys (Alglave, Rivot, et al., 2022; Alglave, Vermard, et al., 2023). These spatio-temporal predictions consist of 132 spatial fields over time.

When the method is univariate (single species) we only retain sole for the analysis. Sole spatio-temporal distribution is strongly structured by reproduction phenology (Arbault, Camus, and Bec, 1986; Alglave, Olmos, et al., 2024). Sole migrates offshore during reproduction from January to March and then returns to nearshore areas for feeding. It is expected that reproduction will be reflected through EOFs. Also, one could constrain EOFs with steady seasonality to better emphasize such process.

In some scenarios, multiple ecological variables must be analyzed together to identify shared dynamics. In such cases, EOFs must be tailored to multivariate data to infer the common patterns shaping these species (Thorson, Scheuerell, et al., 2015). To illustrate this, we consider several species in the Bay of Biscay, including sole, whiting, and squids. These species live in relatively coastal areas and share similar ecological niches (Alglave, Vermard, et al., 2023). Consequently, they are expected to exhibit shared spatio-temporal patterns, such as synchronous reproductive periods from January to March (Moreno et al., 2002;

Arbault, Camus, and Bec, 1986; Houise, 1993). Our objective is to identify and quantify these patterns using multivariate EOFs.

In the following sections, we illustrate various EOFs and EOF-related methods using these case studies.

4 Applying EOFs to the case studies

The EOFs method is applied to satellite NDVI data and the IST-SDM dataset (Figure 1, S1). Respectively, we see an angle in the graph of variance after the second dimension (NDVI) and the fourth dimension (IST-SDM - left, Figure 2). Figure 3 illustrates both the spatial factors (top) and the temporal loadings (bottom) of the first two dimensions for both case studies. The third and fourth dimensions of the IST-SDM case study are presented in Figure S2.

For satellite NDVI data, both the first and second dimensions emphasize seasonality on two different phases. The first dimension is related to the winter/summer seasonality with a positive peak in June and a progressive decrease until February/March. This signal is mainly related to the mountainous and eastern part of France. It is basically the signal that represents (1) the decrease of vegetation in winter (especially in mountainous areas) due to smaller primary production and snow falls and (2) an increase starting in March when temperatures are higher and conditions are favorable for the growth of vegetation.
By contrast, the second dimension emphasizes another phase in seasonality. Temporal loadings are negative from March to June (spring and early summer), which corresponds to high NDVI in the western and northern part of France. This can be related to both rain and higher temperature conducting to high primary production in this area. From July to

November, temporal loadings are positive. NDVI is high in mountains (similar to the first dimension) and low in the western and northern part of France.

For the IST-SDM case study, the first dimension highlights strong seasonality. The loadings exhibit highly positive values in winter (January to February) and strongly negative values during summer (April to August), indicating that soles are concentrated in red/orange areas during winter and in blue areas during summer. The second dimension also emphasizes some seasonality but primarily depicts a long-term decreasing trend. It suggests that fish progressively migrate northward over the period and become more concentrated in Northern red areas. Note that dimension 3 also emphasizes seasonal patterns (Figure S2) related to reproduction and dimension 4 highlights an important hotspot of sole distribution in the south of the Bay of Biscay (2°W - 45.5°N).

EOFs do not disentangle the information between the different dimensions and relevant ecological signals are mixed between the several dimensions of the EOFs. For instance, both the first and second dimensions for the two case studies capture a similar though different seasonal signal (Figure 3). Also, some loadings seem to mix both long term trends and seasonnal trends (see the second EOFs dimension of the IST-SDM case study). Finally, some spatial areas appear to be shared between both spatial factors (mountain areas in the satellite NDVI case study and reproduction areas in the IST-SDM case study).

5 Informing EOFs with an ancillary temporal variable

As mentioned in previous sections, EOFs may not effectively disentangle a single process in each dimension. There are no constraints on the ecological relevance of the spatial factors or temporal loadings, and a single process might be represented across several EOFs dimensions. To address this limitation, it is useful to inform the spatio-temporal dimension

reduction of EOFs with temporal variables that are ecologically meaningfully (also referred to as ancillary variables in the following) to inform the spatio-temporal dimension reduction of EOFs.

This section describes the methodology for extracting EOFs patterns that exhibit the highest correlation with an ancillary temporal variable. This process involves projecting the spatio-temporal data onto the EOFs basis and then performing a CCA between a selected set of EOFs loading factors and the ancillary variable. We first review the CCA method, demonstrate how to adapt it to our specific context, and finally present an application of this technique.

5.1 Basics of Canonical correlation analysis

CCA was first proposed by Hotelling (1992). It is a method used to identify and measure the associations among two sets of variables. Consider two space-time processes $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$ at location x and time t. CCA aims at identifying pairs of spatial basis vectors, $\mathbf{w}_1$ and $\mathbf{w}_2$, for the two processes, $S^{(1)}$ and $S^{(2)}$, that maximize the correlation between their projections onto these vectors. Consider the corresponding observed space-time matrices denoted $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ of size $n_1 \times p$ and $n_2 \times p$, respectively.

The CCA maximizes the following correlation function ρ between $\mathbf{y}_1 = \mathbf{S}^{(1)}\mathbf{w}_1$ and $\mathbf{y}_2 = \mathbf{S}^{(2)}\mathbf{w}_2$:

$$\begin{aligned}\rho &= \frac{\mathbb{E}(\mathbf{y}_1\mathbf{y}_2)}{\sqrt{\mathbb{E}(\mathbf{y}_1)\mathbb{E}(\mathbf{y}_2)}} \\ &= \frac{\mathbb{E}(\mathbf{w}_1^T\mathbf{S}^{(2)\,T}\mathbf{S}^{(2)}\mathbf{w}_2)}{\sqrt{\mathbb{E}(\mathbf{w}_1^T\mathbf{S}^{(2)\,T}\mathbf{S}^{(1)}\mathbf{w}_1)\mathbb{E}(\mathbf{w}_2^T\mathbf{S}^{(1)\,T}\mathbf{S}^{(2)}\mathbf{w}_2)}} \\ &= \frac{\mathbf{w}_1^T\mathbf{C}_{12}\mathbf{w}_2}{\sqrt{\mathbf{w}_1^T\mathbf{C}_{11}\mathbf{w}_1\mathbf{w}_2^T\mathbf{C}_{22}\mathbf{w}_2}}\end{aligned} \tag{6}$$

Here $\mathbf{y}_1$ and $\mathbf{y}_2$ are the temporal variables that arise from the linear combination of $\mathbf{S}^{(1)}\mathbf{w}_1$ and $\mathbf{S}^{(2)}\mathbf{w}_2$ and which correlation coefficient is to be maximised. Then, the spatial basis vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ describe which areas contribute to this correlation in the dataset $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ respectively.

In order to find the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ that maximize ρ, we resort to the method of Lagrange multipliers, leading to a generalized eigenvalue problem (Weenink, 2003).

$$\begin{cases} (\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} - \rho^2)\mathbf{w}_1 = 0 \\ (\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} - \rho^2)\mathbf{w}_2 = 0 \end{cases} \tag{7}$$

Here, the eigenvalues ρ^2 represent the squared canonical correlations, and the eigenvectors $\mathbf{w}_1$ and $\mathbf{w}_2$ are the canonical correlation basis vectors.

5.2 Coupling CCA with EOFs

Note that the two datasets $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ do not necessarily have the same spatial extent, but they must have the same time stepping. Then, to constrain the EOFs with a single ancillary

variable, it is possible to perform a CCA on the matrix $\mathbf{V}$ (Equation 4) composed of the temporal loading factors and an ancillary variable that we denote $\text{An}(t), t \in \{1, \cdots, p\}$. Usually, we only retain the dimensions of the EOFs that capture the process and leave apart the dimension that capture noise (see the rule of thumb in section 2.2.3). In this case, $\mathbf{w}_1$ are coefficients that allow to rotate the temporal loadings as well as the related spatial factors to obtain:

- one variable that closely fit the ancillary variable. This is the combination of loadings that maximise the correlation with the ancillary variable.
- the related spatial factor (or spatial basis vector) that is a combination of the first spatial factors. It will capture as much variance as the first pattern of the EOFs but will not be orthogonal to the following EOFs.

5.3 Illustration

In this section, we inform EOFs of the NDVI satellite data with an ancillary variable representing seasonality. The ancillary variable chosen here is a steady seasonal signal evidencing seasonality with positive peaks in summer (July) and negative peaks in winter (January - red variable in Figure 4). This variable aims at extracting part of the common seasonal signal shared by the two first dimensions of the EOFs (Figure 3). The seasonal signal is parameterized as:

$$\text{An}(t) = A \cdot sin(f \cdot t + D) \tag{8}$$

To have a ancillary variable that match seasonal cycles (opposite peaks in January and July) we set $A = 1$ (the amplitude of the signal), $f = 1/12 \times 2\pi$ the frequency of the signal

components and the delay component $D = 1$. Note that this procedure is not invariant to a change in the phase of the seasonal signal and should be fixed with care.

We then perform a CCA between the first two loading factors of the EOFs (components of the matrix $\mathbf{V}$ in Section 2.2) and the seasonal signal. The aim is to identify spatial pattern of NDVI related to the cycle of seasons (driven by precipitation and temperature in France). The outcomes of this method are presented in Figure 4.

Perform CCA between EOFs results and a seasonnal time series

```
## Construct the seasonnal variable
p <- ncol(S) # number of time steps
ts <- seq(1,p,1) # sequence of time steps
component.strength = 1 # amplitude of the signal
component.freqs = 1/12 # frequency of signal (month^-1)
f.0 = 2 * pi # fundamental frequency (month^-1)
component.delay = 1 # delay of signal components (radians)
# Create the seasonal signal
signal_vec = - component.strength *
    sin(component.freqs * f.0 * ts + component.delay)

## Filter EOFs dimensions and apply CCA
ndim <- 2 # number of dimension to filter in the EOFs
EOFset1 <- E$v[,1:ndim] # filter the loadings
cca_res = stats::cancor(EOFset1, signal_vec) # perform CCA

## Rotate the temporal loadings and spatial factors
## with CCA results
# Rotate the temporal loadings of the EOFs
ccavar1_t <- (as.matrix(EOFset1) %*% cca_res$xcoef)[,1]
# Rotate the spatial factor of the EOFs
ccavar1_x <- as.matrix(E$u[,1:ndim) %*% cca_res$xcoef[,1]
```

The loadings obtained from the CCA emphasize a very good fit to the ancillary variable (R2

= 0.98, Figure 4, top right). By comparison, the first two EOFs loadings have positive, but lower correlation coefficient (R2 = 0.82 and R2 = 0.57). The spatial basis obtained through CCA displays a stronger North/South gradient of NDVI compared with the two first EOFs (Figure 3). NDVI are high in orange/red areas during summer (i.e. the center and mountain part of France) which corresponds to high NDVI in mountains and all the southern part and eastern part of France. In these areas, either ice melt and temperatures become favorable for vegetation growth (mountains), either temperature are high and rains remain frequent which increases vegetation coverage (eastern and south western part of France). In winter, these areas are colder and they can be covered by snow which do not favor vegetation growth.

In blue areas, NDVI is relatively higher during winter. These areas have a cooler climate during winter and strong precipitation which favors vegetation growth.

A similar analysis is performed on the IST-SDM outputs. It allows to better evidence the seasonal cycle of sole reproduction and the offshore-onshore gradients that shapes sole migration (Figure S3)

6 EOFs for multivariate analysis

When studying entire ecosystems or community dynamics, multiple variables need to be analyzed jointly using multivariate modeling approaches (Ovaskainen and Abrego, 2020; Thorson, Ciannelli, and Litzow, 2020). The EOFs framework is flexible and adapts to the multivariate case. It allows to identify the main modes of variability across several variables (or ecosystem compartments) and to detect shared temporal trends and related habitats. In such cases, the underlying EOFs theory remains unchanged (Section 2); only the matrix to be diagonalized $\mathbf{S}'$ is modified.

6.1 The two approaches for conducting multivariate EOFs

Let $v \in \{1, \cdots, s\}$ denotes the index representing the number of variables in the analysis, and let $S^{(v)}(x,t)$ denotes the value of the space-time process for location x, time t, and variable v. We denote by $\mathbf{S'}^{(v)}$ the $n \times p$ matrix of the spatio-temporal variable v.

Two options are available when conducting multivariate EOFs:

1. Construct the matrix by stacking the matrices row-wise. In this case, the matrix is denoted as $\mathbf{S}'^{(row)}_{multi}$ with dimensions $(n \cdot s) \times p$, and it has the following structure:

$$\mathbf{S}'^{(row)}_{multi} = \begin{pmatrix} \mathbf{S}'^{(1)} \\ \mathbf{S}'^{(2)} \\ \vdots \\ \mathbf{S}'^{(s)} \end{pmatrix}$$

2. Alternatively, one can construct the matrix by stacking the matrices column-wise. In this case, the matrix is denoted as $\mathbf{S}'^{(col)}_{multi}$ with dimensions $n \times (p \cdot s)$ and is structured as follows:

$$\mathbf{S}'^{(col)}_{multi} = \begin{pmatrix} \mathbf{S}'^{(1)} & ; & \mathbf{S}'^{(2)} & ; & \cdots & ; & \mathbf{S}'^{(s)} \end{pmatrix}$$

Similar to standard EOFs, SVD enables to infer the spatial factors $\mathbf{u}_i$ and the temporal loadings $\mathbf{v}_i$ (Section 2). However, depending on the data stacking before EOFs, they do not have the same dimensions and do not lead to the same interpretation.

The first case is a **temporally synthetic representation** of the variability in the multivariate data. For each dimension, there is only one single time series of temporal loading and there are as many spatial factors as there are variables. The temporal loadings describe how the different spatial factors evolve over time. This allows to summarize the temporal variation

of all the variables in one single temporal variable. A similar approach is adopted in the VAST package (Thorson, 2019).

The second case, is a **spatially synthetic representation** of the variability in the data. For each dimension, there is only one spatial factor that summarizes the variability of all the variables and as many time series of temporal loadings as there are variables. The loadings of the different variables describe how these variables are related to the spatial factor at a specific time step.

Note finally that in some cases, multivariate EOFs can be applied on a single variable with some time-lag (Wikle, Zammit-Mangion, and Cressie, 2019). In this case the different variables are the lagged version of the spatio-temporal data. This typically allows for a better identification of temporal patterns such as periodicity or long-term trends and is referred to as lag EOFs in this paper. This approach is often referred to as extended EOFs (Weare and Nasstrom, 1982; Wikle, 2002; Dey et al., 2019).

6.2 Illustration

To illustrate multivariate EOFs analysis, we apply the two alternative approaches to the three species in the Bay of Biscay: sole, whiting, and squids. These are coastal species (Alglave, Vermard, et al., 2023) with potentially shared spatio-temporal dynamics *e.g.* same reproduction period (Arbault, Camus, and Bec, 1986; Houise, 1993; Moreno et al., 2002). From an ecological perspective, a critical issue involves identifying shared spatial or temporal dynamics among different species (Thorson, Scheuerell, et al., 2015; Ovaskainen and Abrego, 2020; Thorson, Arimitsu, et al., 2021). Additionally, one can seek to identify combinations of 'areas x seasons' that can be interpreted as multispecies functional habitats (Delage and Le Pape, 2016). To do this, we combine the EOFs (both representations)

with a clustering analysis (HCPC) to identify clusters of locations and time steps that can be interpreted as habitats and seasons from a multispecies perspective. For a detailed description of the clustering analysis performed on the EOFs results, see S.1.6 and Alglave, Olmos, et al. (2024).

The temporally synthetic multivariate EOFs (EOFs on $\mathbf{S}'^{(row)}_{multi}$) provides seasonal temporal loadings that represent the reproduction phenology for all the species (as mentioned in Section 3). We select the first two dimensions (Figure 2). For sole, the reproduction areas are generally consistent with those identified when conducting univariate EOFs (Figure 5). For squids and whitings, reproduction areas are observed along the Landes coast (45°N - 1.5°W), on the Vendée coast (46.5°N - 2°W), and in the north of the Bay of Biscay (47.5°N - 3°W). The clustering confirms these results (Figure S10) with an additional potential reproduction area off the continental shelf on the slope (3°W - 45°N).

On the other hand, the spatially synthetic multivariate EOFs (EOFs on $\mathbf{S}'^{(col)}_{multi}$) provides one single spatial pattern per dimension and as many temporal loadings as variables. An angle appears after the fourth dimension (Figure 2). Interestingly, the first spatial factor gathers the main reproduction areas for sole and squids that were previously identified (Figure 6 and S7). Also for all three species a summer area appears on the coast consistently with the clustering analysis (Figure S13). The second dimension emphasizes an overall increase in the loadings, illustrating a general drift towards the North of the Bay of Biscay and in the offshore areas (Figure S8). The loadings are correlated among species, highlighting that they share common temporal dynamics relatively to the spatial factors of these dimensions (Figure 6). The third dimension is mainly characterized by whiting reproduction areas and the fourth dimension by sole reproduction areas (Figure S9).

Illustration of the lag EOFs is realised on the satellite NDVI data and is presented in S.1.4 as a complement of the multispecies analysis. Compared with standard EOFs, they allow to

identify an additional periodic signal with a 6-months phase related to an alternance of high growth of vegetation in spring and autumn and lower growth in summer and winter.

7 Discussion

Towards a wider use of EOFs in ecology

EOFs have tremendous potential for ecology. As massive and diverse datasets become available through various data sources (such as satellite data, IST-SDM, and biogeochemical models), these decomposition methods will play a growing role in (1) reducing data dimensionality and (2) extracting ecologically relevant and interpretable information.

While these methods have been frequently used in marine ecology (Norton and Mason, 2005; Petitgas, Doray, et al., 2014; Thorson, Cheng, et al., 2020), they have remained underutilized in terrestrial ecology. In this paper, we demonstrate their potential for both terrestrial and marine systems. We argue that these methods should be systematically applied whenever the time span of the data necessitates post-processing to summarize the information.

Accounting for spatio-temporal correlation in EOFs

One might argue that EOFs do not explicitly account for spatial or temporal correlations and then are not per se a spatio-temporal method. It is true that EOFs is basically a PCA representation of spatio-temporal data; the spatio-temporal aspects of the method comes from the characteristic of the data rather than from the mathematical formulation of the method per se.

In a spatial context, methods have been developed to better handle spatial correlations in data and to propose multiscale decompositions of ecological data (Dray et al., 2012; P. Legendre and L. Legendre, 2012b). P. Legendre and L. Legendre (2012a) provides an

overview of 'spatial eigenfunction analysis' (e.g., Moran's eigenvector maps, asymmetric eigenvector maps, multiscale ordination), which estimate sets of eigenvectors based on spatial configuration matrices. These eigenvectors can be used as predictors in linear models, canonical analysis, or redundancy analysis. They help identify different spatial scales of variation in ecological data and the related spatial patterns. These methods could potentially be extended to spatio-temporal data to identify distinct spatio-temporal scales that structure the ecological processes.

More recently, Petitgas, Renard, et al. (2020) and Bez, Renard, and Ahmed-Babou (2023) have proposed a method (Empirical Orthogonal Maps - EOM) based on EOFs that accounts for spatial correlation by rotating the spatial factors to provide spatially decorrelated patterns. EOM provide maps with stronger orthogonality constraints to better separate the information carried by different spatial factors. Alternatively, spatial correlations can be represented as spatial latent variables, and EOFs can be applied to those within a hierarchical model (Thorson, 2019).

Although these methods are sensitive to the choice of the distance matrix or the spatial lags used in the analysis, they represent promising avenues for enhancing EOFs to explicitly account for spatial and temporal correlation.

Informing EOFs with ancillary variables for a better ecological interpretation

A significant drawback of EOFs is their limited interpretability regarding the processes underlying the spatio-temporal field particularly in the dimensions up to the second dimension (Mestas-Nuñez, 2000; Monahan et al., 2009). This issue can be attributed to the orthogonal constraints, which do not have per se an ecological meaning. Under Gaussian assumptions, orthogonal constraints allow for the extraction of independent information across different dimensions and facilitate the separation of information along the axes. However, in ecology,

(1) processes are frequently non-Gaussian and (2) can be highly correlated, non-linear, and non-stationary (Ciannelli, Bartolino, and Chan, 2012). Consequently, standard EOFs may not always provide the best ecological representation of the data.

Here drawing on CCA theory (Hotelling, 1992), we propose an approach that enhances the interpretability of the spatial factors by informing the EOFs with an ecological ancillary temporal variable. A key consideration is the type of ecological ancillary variable that can be used to constrain the EOFs. Ideally, any temporal variable with the same temporal extent as the spatio-temporal dataset can be used, provided it has ecological relevance to the ecological question under study. For example, many studies have explored the relationship between EOFs and oceanographic processes, such as the cold pool extent in the Bering Sea (Thorson, Ciannelli, and Litzow, 2020; Thorson, Arimitsu, et al., 2021). Constraining the EOFs with the cold pool extent would likely improve the interpretability of the spatial patterns. In other contexts, ancillary variables could include other oceanographic processes, such as upwelling intensity (Sydeman et al., 2014), or any covariates (or sets of covariates) that drive the spatio-temporal variability of populations and ecosystems (Petitgas, Doray, et al., 2014).

EOFs for community and ecosystem analysis

The use of EOFs for multivariate data has gained significant importance over the past decade. Initially, research focused on synthesizing information from temperature-salinity diagrams to reduce the complexity of data representation (Fukumori and Wunsch, 1991; Sparnocchia, Pinardi, and Demirov, 2003). More recently, EOFs have been applied to study ecosystem structure and community dynamics (Woillez et al., 2010; Petitgas, Huret, et al., 2018; Grandremy et al., 2023) and to forecast these communities and ecosystems (Thorson, Arimitsu, et al., 2021).

Using three coastal fish species in the Bay of Biscay, we demonstrated the ability of EOFs to (1) identify shared spatial and temporal patterns among species, (2) evaluate correlations among species, and (3) identify seasons and functional zones common to these species. Two kind of representation are presented in this paper: a spatially synthetic representation and a temporally synthetic presentation. The latter approach aligns with the VAST framework (Thorson, Ciannelli, and Litzow, 2020; Thorson, 2019), while the former is less commonly seen in the literature but remains valuable for identifying shared spatial patterns across multiple variables or species.

Developing multivariate EOFs that are both spatially and temporally synthetic is crucial to enhance the interpretability of results. EOFs can be applied to any set of variables (species or ecological variable), potentially encompassing many more than three, to represent several communities or entire ecosystems. The methodological steps and interpretations presented here would remain applicable. However, as the number of variables increases, either the number of variables contained in the temporal loadings or in the spatial factors increases, leading to results that are difficult to interpret. For example, recent studies by Thorson, Ciannelli, and Litzow (2020) and Thorson, Arimitsu, et al. (2021) proposed multivariate EOFs to investigate ecosystem variability by including respectively seven and eighteen variables. In this case, it results in a set of seven and eighteen maps per EOFs dimension and visual inspection does not allow to have a meaningful and synthetic view of the information. New methods based on multidimensional array SVD could be employed to develop multivariate spatio-temporal EOFs with more synthetic representations of the temporal and spatial components (Bi et al., 2021). In particular, future research could incorporate multiple loadings, representing separate the expansion across spatial, temporal, and multivariate dimensions.

Non-stationary, non-linear EOFs for spatio-temporal analysis under climate change

Finally, some important hypotheses underlie EOFs that can be an important drawback in a context of climate change, namely stationarity and linearity (Hannachi, Jolliffe, and Stephenson, 2007). Indeed, the spatial patterns identified through EOFs remain unchanged throughout the period. The spatial EOFs factors may decrease or increase in intensity through the linear combination with the temporal loadings, but they will remain the same over the period (Alglave, Olmos, et al., 2024). Consequently, EOFs can not be used to describe propagating patterns such as range expansion or contraction as is common in species redistribution (Cheung et al., 2009; Melles et al., 2011; Scheele et al., 2017). Also, as in standard PCA, all non-linear processes will not be detected and captured through EOFs while it is largely recognized that ecosystems are strongly structured by non-linear relationships and sometimes switching dynamics (Levin, 1998; Scheffer et al., 2001). More generally speaking, handling non-stationarity in a changing environment is a key challenge for statistical ecology (Litzow et al., 2019; Astigarraga et al., 2020; Bueno de Mesquita et al., 2021). Although some techniques, such as Hilbert complex EOFs or non-linear PCA (Bueso, Piles, and Camps-Valls, 2020; Esquivel and Messina, 2008), allow to handle non-linear relationship and to evidence propagating patterns, these have been scarcely used in practice to investigate the effect of climate change on ecosystems. Developing and applying approaches that can handle such non-linearities and non-stationarity is critical in analysing the effect of climate change on ecosystems; it constitues an open research avenue for future study of threatened ecosystems in a perspective of conservation and adaptation to climate change (Sala et al., 2000; Parmesan, 2006; Grebmeier et al., 2006).

References

Alglave, Baptiste, Maxime Olmos, et al. (2024). "Investigating fish reproduction phenology and essential habitats by identifying the main spatio-temporal patterns of fish distribution". In: ICES Journal of Marine Science, fsae099.

Alglave, Baptiste, Etienne Rivot, et al. (2022). "Combining scientific survey and commercial catch data to map fish distribution". In: ICES Journal of Marine Science 79.4, pp. 1133–1149.

Alglave, Baptiste, Youen Vermard, et al. (2023). "Identifying mature fish aggregation areas during spawning season by combining catch declarations and scientific survey data". In: Canadian Journal of Fisheries and Aquatic Sciences 80.5, pp. 808–824.

Arbault, Par Suzanne, P Camus, and C le Bec (1986). "Estimation du stock de sole (Solea vulgaris, Quensel 1806) dans le Golfe de Gascogne à partir de la production d'oeufs". In: Journal of applied ichthyology 2.4, pp. 145–156.

Astigarraga, Julen et al. (2020). "Evidence of non-stationary relationships between climate and forest responses: Increased sensitivity to climate change in Iberian forests". In: Global Change Biology 26.9, pp. 5063–5076.

Berkelhammer, Max (2019). "Synchronous modes of terrestrial and marine productivity in the North Pacific". In: Frontiers in Earth Science 7, p. 73.

Bez, Nicolas, Didier Renard, and Dedah Ahmed-Babou (2023). "Empirical Orthogonal Maps (EOM) and Principal Spatial Patterns: Illustration for Octopus Distribution Off Mauritania Over the Period 1987–2017". In: Mathematical Geosciences 55.1, pp. 113–128.

Bi, Xuan et al. (2021). "Tensors in statistics". In: Annual review of statistics and its application 8.1, pp. 345–368.

Bueno de Mesquita, Clifton P et al. (2021). "Taking climate change into account: Non-stationarity in climate drivers of ecological response". In: Journal of Ecology 109.3, pp. 1491–1500.

Bueso, Diego, Maria Piles, and Gustau Camps-Valls (2020). "Nonlinear PCA for spatio-temporal analysis of Earth observation data". In: IEEE Transactions on Geoscience and Remote Sensing 58.8, pp. 5752–5763.

Busetto, Lorenzo and Luigi Ranghetti (2016). "MODIStsp: an R package for preprocessing of MODIS Land Products time series". In: Computers & Geosciences 97, pp. 40–48. ISSN: 0098-3004. DOI: 10.1016/j.cageo.2016.08.020. URL: https://github.com/ropensci/MODIStsp.

Cheung, William WL et al. (2009). "Projecting global marine biodiversity impacts under climate change scenarios". In: Fish and fisheries 10.3, pp. 235–251.

Ciannelli, Lorenzo, Valerio Bartolino, and Kung-Sik Chan (2012). "Non-additive and non-stationary properties in the spatial distribution of a large marine fish population". In: Proceedings of the Royal Society B: Biological Sciences 279.1743, pp. 3635–3642.

Delage, Nicolas and Olivier Le Pape (2016). Inventaire des zones fonctionnelles pour les ressources halieutiques dans les e Rapport de recherche. Rennes: Pôle halieutique AGROCAMPUS OUEST, p. 36.

Dey, A et al. (2019). "An operational tracking method for the MJO using extended empirical orthogonal functions". In: Pure and Applied Geophysics 176, pp. 2697–2717.

Di Lorenzo, Emanuele et al. (2008). "North Pacific Gyre Oscillation links ocean climate and ecosystem change". In: Geophysical research letters 35.8.

Dray, Stéphane et al. (2012). "Community ecology in the age of multivariate multiscale spatial analysis". In: Ecological Monographs 82.3, pp. 257–275.

Eastman, J Ronald et al. (2013). "Global trends in seasonality of normalized difference vegetation index (NDVI), 1982–2011". In: Remote Sensing 5.10, pp. 4799–4818.

612 Eisfelder, Christina et al. (2023). “Seasonal Vegetation Trends for Europe over 30 Years from
613 a Novel Normalised Difference Vegetation Index (NDVI) Time-Series—The TIMELINE
614 NDVI Product”. In: Remote Sensing 15.14, p. 3616.
615 Esquivel, P and AR Messina (2008). “Complex empirical orthogonal function analysis of
616 wide-area system dynamics”. In: 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of
617 IEEE, pp. 1–7.
618 Fukumori, Ichiro and Carl Wunsch (1991). “Efficient representation of the North At-
619 lantic hydrographic and chemical distributions”. In: Progress in Oceanography 27.1-2,
620 pp. 111–195.
621 Fukuoka, A (1951). “The central meteorological observatory, a study on 10-day forecast (a
622 synthetic report)”. In: Geophysical Magazine 22.3, pp. 177–208.
623 Garrigues, Sébastien, Denis Allard, and Frédéric Baret (2008). “Modeling tem-
624 poral changes in surface spatial heterogeneity over an agricultural site”. In:
625 Remote Sensing of Environment 112.2, pp. 588–602.
626 Gedalof, Ze’ev, David L Peterson, and Nathan J Mantua (2005). “Atmospheric, climatic,
627 and ecological controls on extreme wildfire years in the northwestern United States”. In:
628 Ecological Applications 15.1, pp. 154–174.
629 Grandremy, Nina et al. (2023). “Hydrology and small pelagic fish drive the spatio–temporal
630 dynamics of springtime zooplankton assemblages over the Bay of Biscay continental
631 shelf”. In: Progress in Oceanography 210, p. 102949.
632 Grebmeier, Jacqueline M et al. (2006). “A major ecosystem shift in the northern Bering
633 Sea”. In: Science 311.5766, pp. 1461–1464.
634 Hannachi, Abdel, Ian T Jolliffe, and David B Stephenson (2007). “Empirical or-
635 thogonal functions and related techniques in atmospheric science: A review”. In:
636 International Journal of Climatology: A Journal of the Royal Meteorological Society
637 27.9, pp. 1119–1152.

Hotelling, Harold (1992). “Relations between two sets of variates”. In: Breakthroughs in statistics: methodology and distribution. Springer, pp. 162–190.

Houise, Christophe (1993). “Etude de la population du merlan (Merlangius merlangus L.) du golfe de Gascogne”. In.

Hurrell, James W and Clara Deser (2010). “North Atlantic climate variability: the role of the North Atlantic Oscillation”. In: Journal of marine systems 79.3-4, pp. 231–244.

Iida, Takahiro and Sei-Ichi Saitoh (2007). “Temporal and spatial variability of chlorophyll concentrations in the Bering Sea using empirical orthogonal function (EOF) analysis of remote sensing data”. In: Deep Sea Research Part II: Topical Studies in Oceanography 54.23-26, pp. 2657–2671.

Isaac, Nick JB et al. (2020). “Data integration for large-scale models of species distributions”. In: Trends in ecology & evolution 35.1, pp. 56–67.

Le Mezo, Priscilla et al. (2016). “Natural variability of marine ecosystems inferred from a coupled climate to ecosystem simulation”. In: Journal of Marine Systems 153, pp. 55–66.

Legendre, Pierre and Louis Legendre (2012a). “Multiscale analysis: spatial eigenfunctions”. In: Developments in environmental modelling. Vol. 24. Elsevier, pp. 859–906.

— (2012b). Numerical ecology. Vol. 24. Elsevier.

Levin, Simon A (1998). “Ecosystems and the biosphere as complex adaptive systems”. In: Ecosystems 1, pp. 431–436.

Lindegren, Martin et al. (2022). “A spatial statistical approach for identifying population structuring of marine fish species: European sprat as a case study”. In: ICES Journal of Marine Science 79.2, pp. 423–434.

Litzow, Michael A et al. (2019). Nonstationary environmental and community relationships in the North Pacific Ocean.

Lorenz, Edward N (1956). Empirical orthogonal functions and statistical weather prediction. Vol. 1. Massachusetts Institute of Technology, Department of Meteorology Cambridge.

Martin, Beatriz et al. (2020). “Citizen science for predicting spatio-temporal patterns in seabird abundance during migration”. In: Plos one 15.8, e0236631.

Melles, SJ et al. (2011). “Expanding northward: influence of climate change, forest connectivity, and population processes on a threatened species’ range shift”. In: Global Change Biology 17.1, pp. 17–31.

Mestas-Nuñez, Alberto M (2000). “Orthogonality properties of rotated empirical modes”. In: International Journal of Climatology: A Journal of the Royal Meteorological Society 20.12, pp. 1509–1516.

MODIS (2021). MOD13Q1 V6. Accessed: 2024-07-11. URL: https://lpdaac.usgs.gov/products/mod13q1v006/.

Monahan, A. H. et al. (2009). “Empirical Orthogonal Functions: The Medium is the Message”. In: Journal of Climate 22.24, pp. 6501–6514. DOI: 10.1175/2009JCLI3062.1. URL: https://journals.ametsoc.org/view/journals/clim/22/24/2009jcli3062.1.xml.

Moreno, A et al. (2002). “Biological variation of Loligo vulgaris (Cephalopoda: Loliginidae) in the eastern Atlantic and Mediterranean”. In: Bulletin of Marine Science 71.1, pp. 515–534.

Nathan, Ran et al. (2022). “Big-data approaches lead to an increased understanding of the ecology of animal movement”. In: Science 375.6582, eabg1780.

Norton, JERROLD G and JANET E Mason (2005). “Relationship of California sardine (Sardinops sagax) abundance to climate-scale ecological changes in the California Current system”. In: California Cooperative Oceanic Fisheries Investigations Report 46, p. 83.

Obukhov, Alexander M (1947). “Statistically homogeneous fields on a sphere”. In: Usp. Mat. Nauk 2.2, pp. 196–198.

689 Ottersen, Geir et al. (2001). “Ecological effects of the North Atlantic oscillation”. In:
690 Oecologia 128, pp. 1–14.
691 Ovaskainen, Otso and Nerea Abrego (2020). Joint species distribution modelling: With applications in R.
692 Cambridge University Press.
693 Parmesan, Camille (2006). “Ecological and evolutionary responses to recent climate change”.
694 In: Annu. Rev. Ecol. Evol. Syst. 37.1, pp. 637–669.
695 Petitgas, Pierre (2008). “Fish habitat mapping with empirical orthogonal functions”. In:
696 International Council for the Exploration of the Sea.
697 Petitgas, Pierre, Mathieu Doray, et al. (2014). “Modelling the variability in fish spatial
698 distributions over time with empirical orthogonal functions: anchovy in the Bay of
699 Biscay”. In: ICES Journal of Marine Science 71.9, pp. 2379–2389.
700 Petitgas, Pierre, Martin Huret, et al. (2018). “Ecosystem spatial structure revealed by
701 integrated survey data”. In: Progress in Oceanography 166, pp. 189–198.
702 Petitgas, Pierre, Jacques Masse, et al. (2006). “Hydro-plankton characteristics and their
703 relationship with sardine and anchovy distributions on the French shelf of the Bay of
704 Biscay”. In: Scientia Marina 70.1, pp. 161–172.
705 Petitgas, Pierre, Didier Renard, et al. (2020). “Analysing temporal variability in spatial
706 distributions using min–max autocorrelation factors: sardine eggs in the bay of biscay”.
707 In: Mathematical Geosciences 52, pp. 337–354.
708 Piao, Shilong et al. (2003). “Interannual variations of monthly and seasonal nor-
709 malized difference vegetation index (NDVI) in China from 1982 to 1999”. In:
710 Journal of Geophysical Research: Atmospheres 108.D14.
711 Poggi, Sylvain et al. (2021). “How can models foster the transition towards future agricultural
712 landscapes?” In: Advances in Ecological Research. Vol. 64. Elsevier, pp. 305–368.
713 Roundy, P. E. (2015). “On the Interpretation of EOF Analysis of ENSO, Atmospheric Kelvin
714 Waves, and the MJO”. In: Journal of Climate 28.3, pp. 1148–1165. DOI: 10.1175/JCLI-

715 D-14-00398.1. URL: https://journals.ametsoc.org/view/journals/clim/28/3/
716 jcli-d-14-00398.1.xml.
717 Saeed, Sajjad, Fred Kucharski, and Mansour Almazroui (2023). “Impacts of mid-latitude
718 circulation on winter temperature variability in the Arabian Peninsula: the explicit role
719 of NAO”. In: Climate Dynamics 60.1, pp. 147–164.
720 Sala, Osvaldo E et al. (2000). “Global biodiversity scenarios for the year 2100”. In: science
721 287.5459, pp. 1770–1774.
722 Scheele, Ben C et al. (2017). “Niche contractions in declining species: mechanisms and
723 consequences”. In: Trends in Ecology & Evolution 32.5, pp. 346–355.
724 Scheffer, Marten et al. (2001). “Catastrophic shifts in ecosystems”. In: Nature 413.6856,
725 pp. 591–596.
726 Schrum, Corinna, Mike St John, and Irina Alekseeva (2006). “ECOSMO, a coupled
727 ecosystem model of the North Sea and Baltic Sea: Part II. Spatial-seasonal characteristics
728 in the North Sea as revealed by EOF analysis”. In: Journal of Marine Systems 61.1-2,
729 pp. 100–113.
730 Seaton, Fiona M, Susan G Jarvis, and Peter A Henrys (2024). “Spatio-
731 temporal data integration for species distribution modelling in R-INLA”. In:
732 Methods in Ecology and Evolution.
733 Sparnocchia, S, N Pinardi, and E Demirov (2003). “Multivariate Empirical Orthogonal
734 Function analysis of the upper thermocline structure of the Mediterranean Sea from
735 observations and model simulations”. In: Annales Geophysicae. Vol. 21. 1. Copernicus
736 GmbH, pp. 167–187.
737 Sullivan, Brian L et al. (2014). “The eBird enterprise: An integrated approach to development
738 and application of citizen science”. In: Biological conservation 169, pp. 31–40.
739 Sydeman, WJ et al. (2014). “Climate change and wind intensification in coastal upwelling
740 ecosystems”. In: Science 345.6192, pp. 77–80.

741 Thorson, James T (2019). “Guidance for decisions using the Vector Autoregressive Spatio-
742 Temporal (VAST) package in stock, ecosystem, habitat and climate assessments”. In:
743 Fisheries Research 210, pp. 143–161.
744 Thorson, James T, Sean C Anderson, et al. (2024). “tinyVAST: R package with an expressive
745 interface to specify lagged and simultaneous effects in multivariate spatio-temporal
746 models”. In: arXiv preprint arXiv:2401.10193.
747 Thorson, James T, Mayumi L Arimitsu, et al. (2021). “Forecasting community reassembly
748 using climate-linked spatio-temporal ecosystem models”. In: Ecography 44.4, pp. 612–
749 625.
750 Thorson, James T, Wei Cheng, et al. (2020). “Empirical orthogonal function regression:
751 Linking population biology to spatial varying environmental conditions using climate
752 projections”. In: Global Change Biology 26.8, pp. 4638–4649.
753 Thorson, James T, Lorenzo Ciannelli, and Michael A Litzow (2020). “Defining indices of
754 ecosystem variability using biological samples of fish communities: a generalization of
755 empirical orthogonal functions”. In: Progress in Oceanography 181, p. 102244.
756 Thorson, James T, Mark D Scheuerell, et al. (2015). “Spatial factor analysis: a new
757 tool for estimating joint species distributions and correlations in species range”. In:
758 Methods in Ecology and Evolution 6.6, pp. 627–637.
759 Van Audenhaege, Loıc et al. (2022). “Long-term monitoring reveals unprecedented stability
760 of a vent mussel assemblage on the Mid-Atlantic Ridge”. In: Progress in Oceanography
761 204, p. 102791.
762 Vicente-Serrano, Sergio M et al. (2013). “Response of vegetation to drought time-scales
763 across global land biomes”. In: Proceedings of the National Academy of Sciences
764 110.1, pp. 52–57.
765 Weare, Bryan C and John S Nasstrom (1982). “Examples of extended empirical orthogonal
766 function analyses”. In: Monthly Weather Review 110.6, pp. 481–485.

Weenink, David (2003). “Canonical correlation analysis”. In: Proceedings of the Institute of Phonetic Sciences of the Univ
Vol. 25. University of Amsterdam Amsterdam, pp. 81–99.
Wikle, Christopher K (2002). Spatio-temporal methods in climatology.
Wikle, Christopher K, Andrew Zammit-Mangion, and Noel Cressie (2019). Spatio-temporal statistics with R. Chapman and Hall/CRC.
Woillez, Mathieu et al. (2010). “Statistical monitoring of spatial patterns of environmental indices for integrated ecosystem assessment: Application to the Bay of Biscay pelagic zone”. In: Progress in Oceanography 87.1-4, pp. 83–93.
Wu, Qiong et al. (2023). “Spatiotemporal variations of water conservation function based on EOF analysis at multi time scales under different ecosystems of Heihe River Basin”. In: Journal of Environmental Management 325, p. 116532.
Zipkin, Elise F et al. (2021). “Addressing data integration challenges to link ecological processes across scales”. In: Frontiers in Ecology and the Environment 19.1, pp. 30–38.

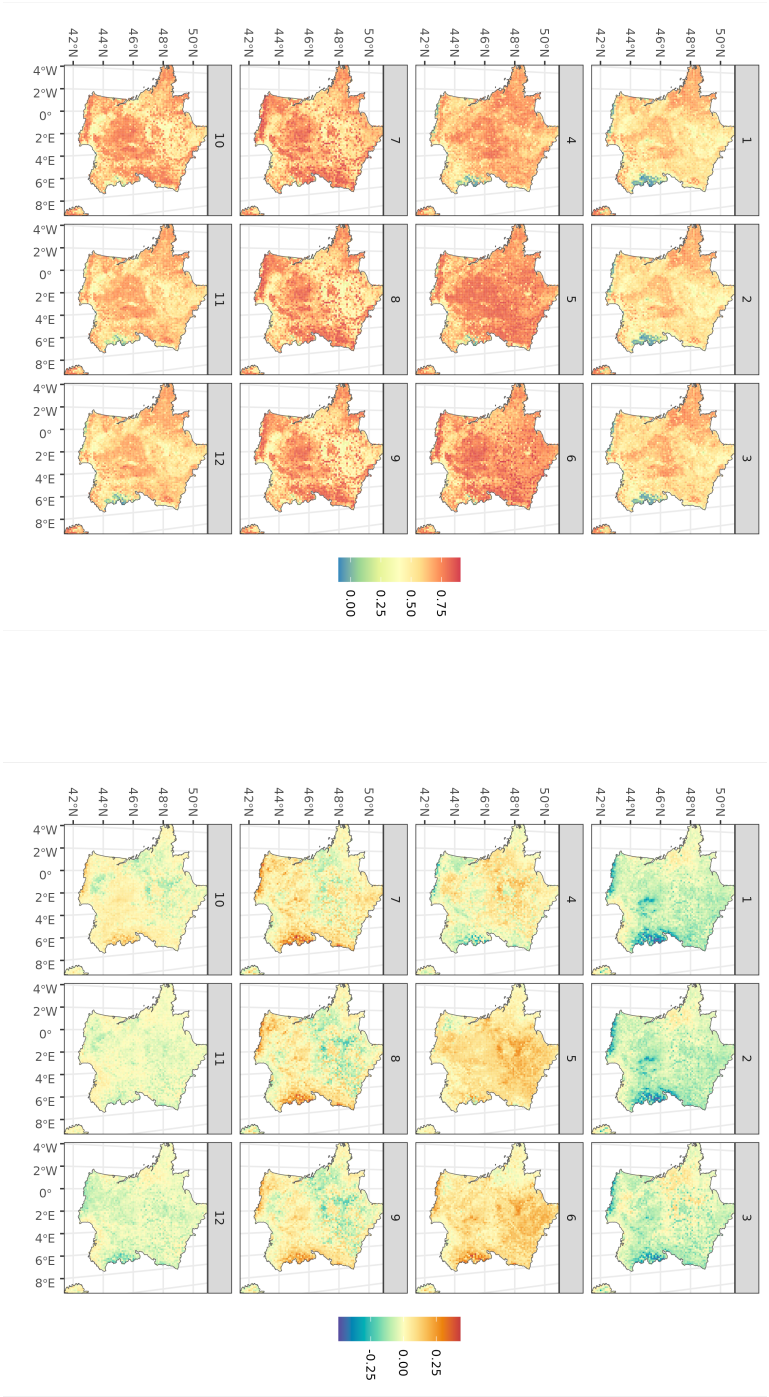

Fig. 1: NDVI data. (left) Monthly spatial data of NDVI $S(x,t)$. (right) Monthly anomalies of the spatial predictions $S^*(x,t)$. Each panel corresponds to the average predictions or anomalies for a month over the period 2000 - 2023.

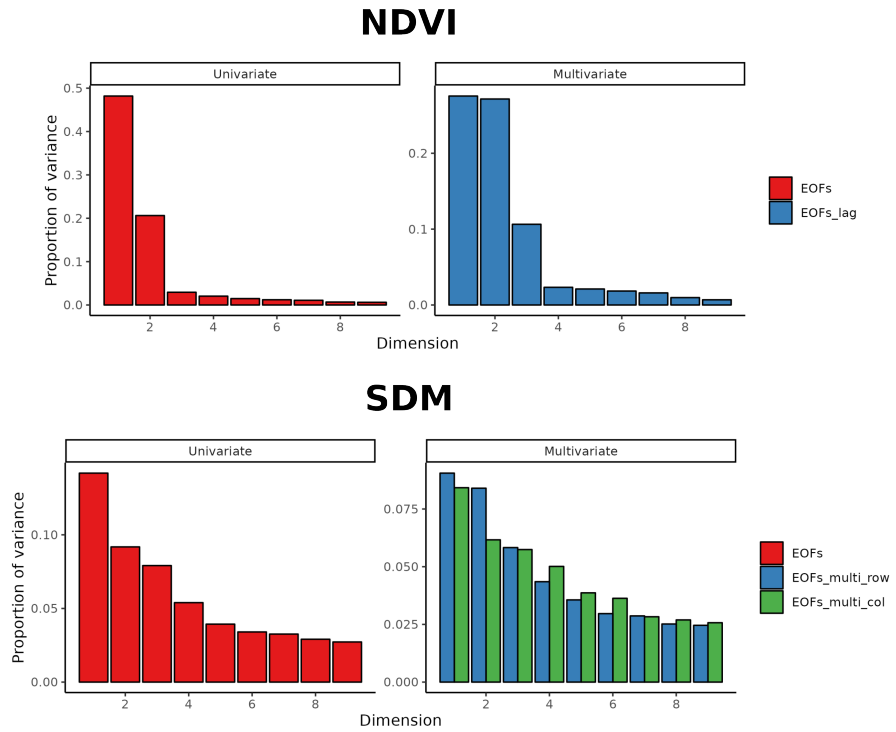

Fig. 2: Proportion of variance explained by each method for each case study. Top: NDVI satellite data, bottom: IST-SDM predictions for sole. (Left) Univariate analysis. (Right) Multivariate analysis. EOFs_lag is the lagged version of EOFs. EOFs_multi_row is the multivariate EOFs conducted on $\mathbf{S}_{multi}^{(row)}$ (temporally synthetic representation). EOFs_multi_col is the multivariate EOFs conducted on $\mathbf{S}_{multi}^{(col)}$ (spatially synthetic representation).

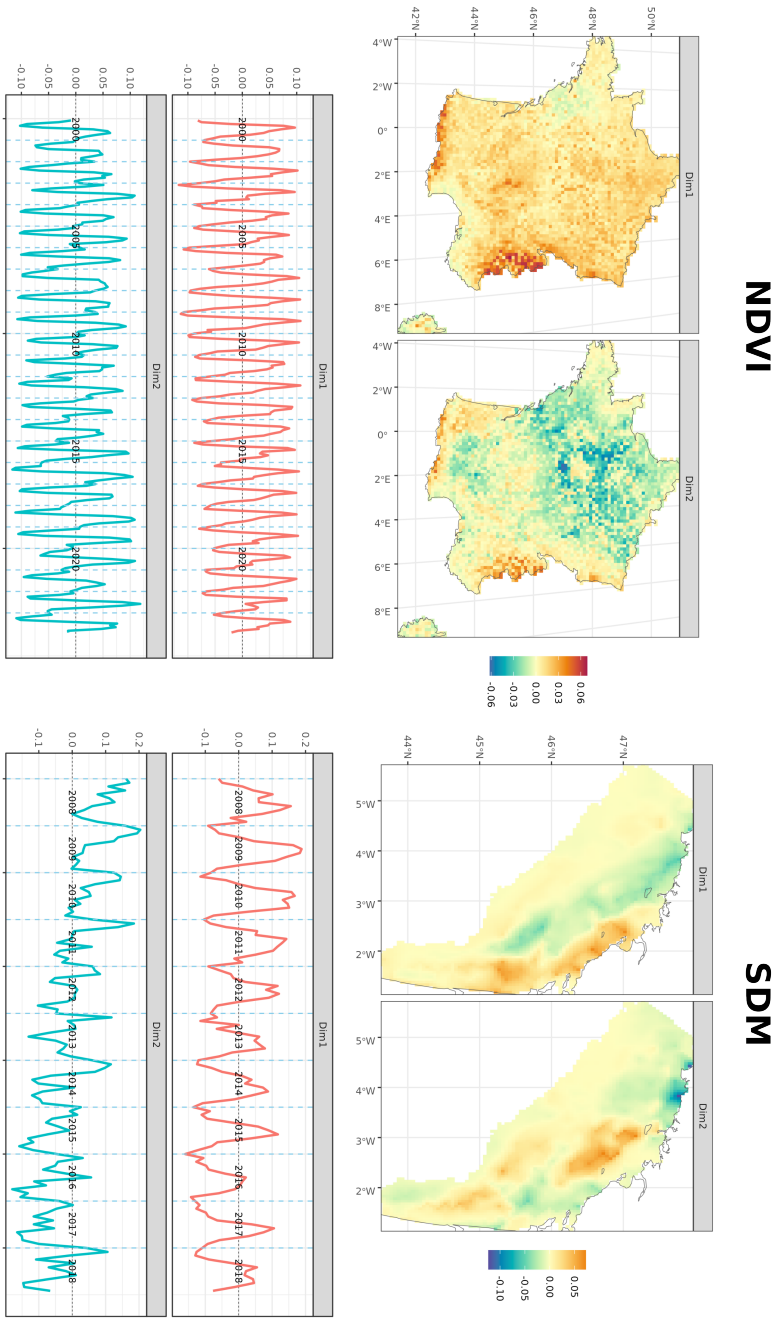

Fig. 3: EOFs results for satellite NDVI data (left) and IST-SDM predictions (right). (Top) Spatial factors for the two first dimensions of the EOFs. (Bottom) Loadings for the two first dimensions of the EOFs. Blue dashed vertical lines correspond to the month of January for each year.

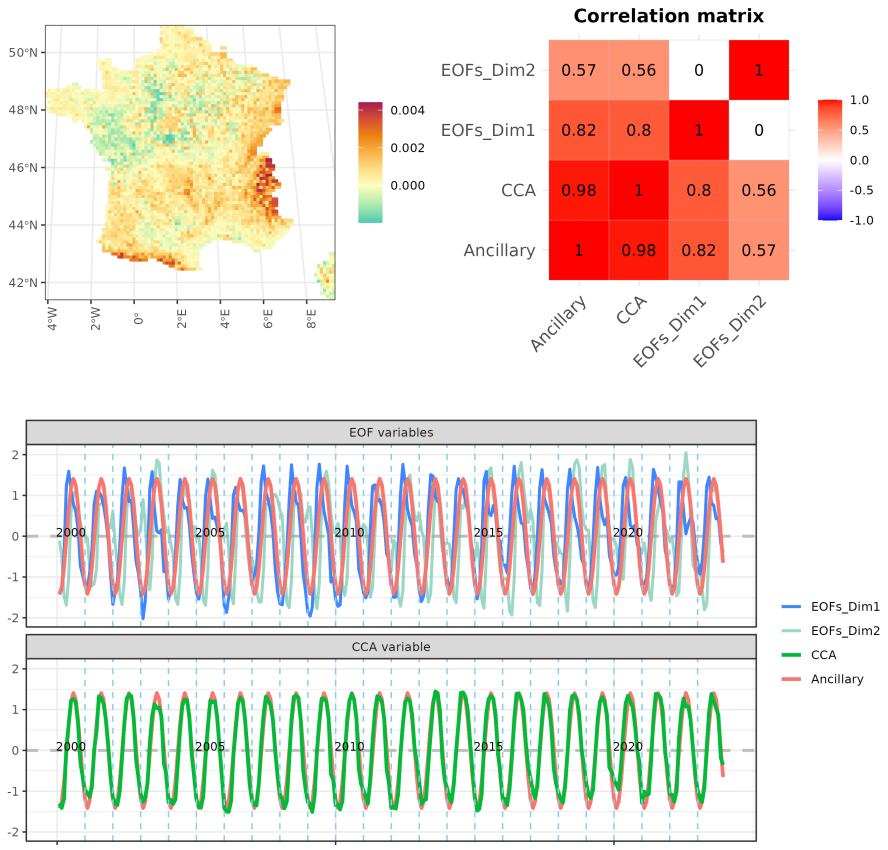

Fig. 4: Satellite NDVI data. Results of the canonical correlation analysis. (Top left) Spatial basis vector that maximize the correlation between the temporal variables $\mathbf{y}_1$ and $\mathbf{y}_2$. (Top right) Correlation matrix between the two first loading factors of the EOFs, the CCA $\mathbf{y}_1$, and the ancillary variable. (Bottom) Comparison of the EOFs loadings with the ecological ancillary variable (showing that both variables capture some seasonal pattern), and the CCA variable with the ecological ancillary variable (showing that the CCA now captures seasonality in a single variable). These time series are standardized. Blue dashed vertical lines correspond to the month of January for each year.

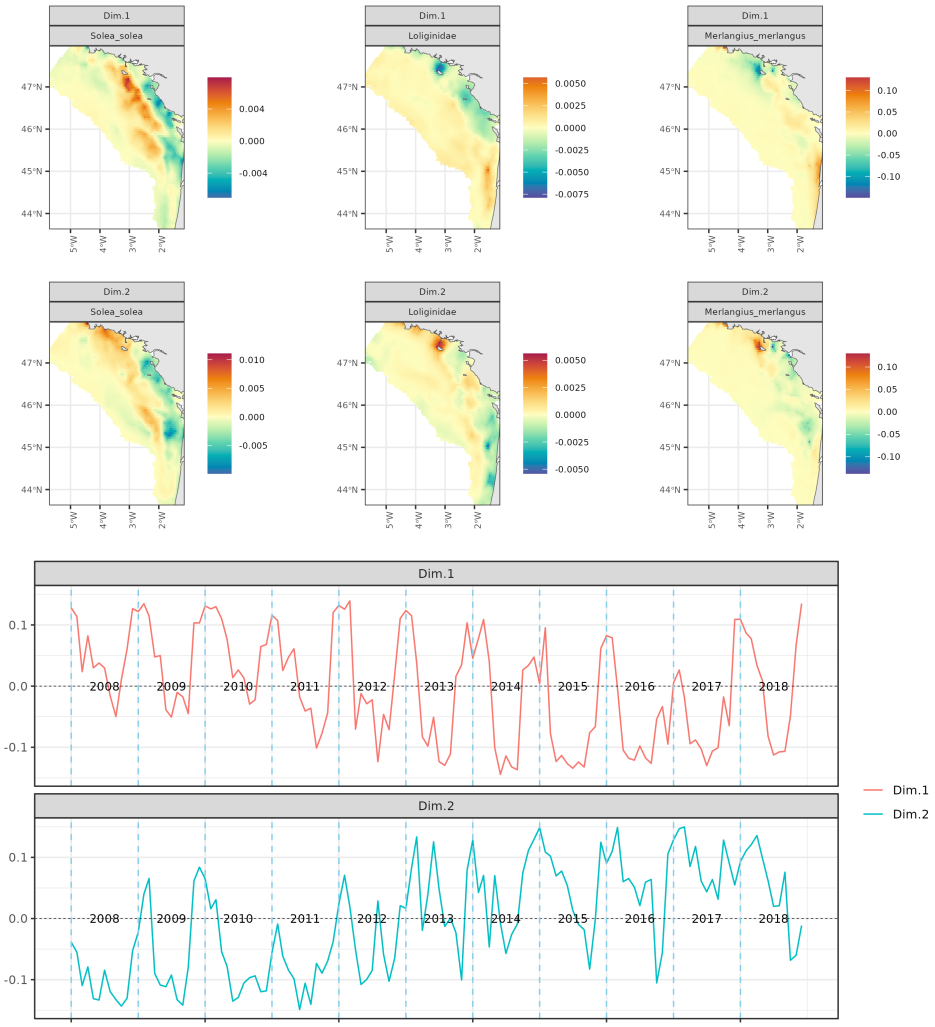

Fig. 5: IST-SDM predictions. Results for the multivariate temporally synthetic EOFs (EOFs on $\mathbf{S}'^{(row)}_{multi}$). (Top) Factor maps for each species and first two dimensions. (Bottom) Loadings for the first two dimensions. For each dimension, loadings represent when the several species of the analysis follow their related spatial distribution. Blue dashed vertical lines correspond to the month of January for each year. The temporal loadings emphasize the seasonality of the reproduction for each three species. Orange areas in both dimensions correspond to the reproduction grounds.

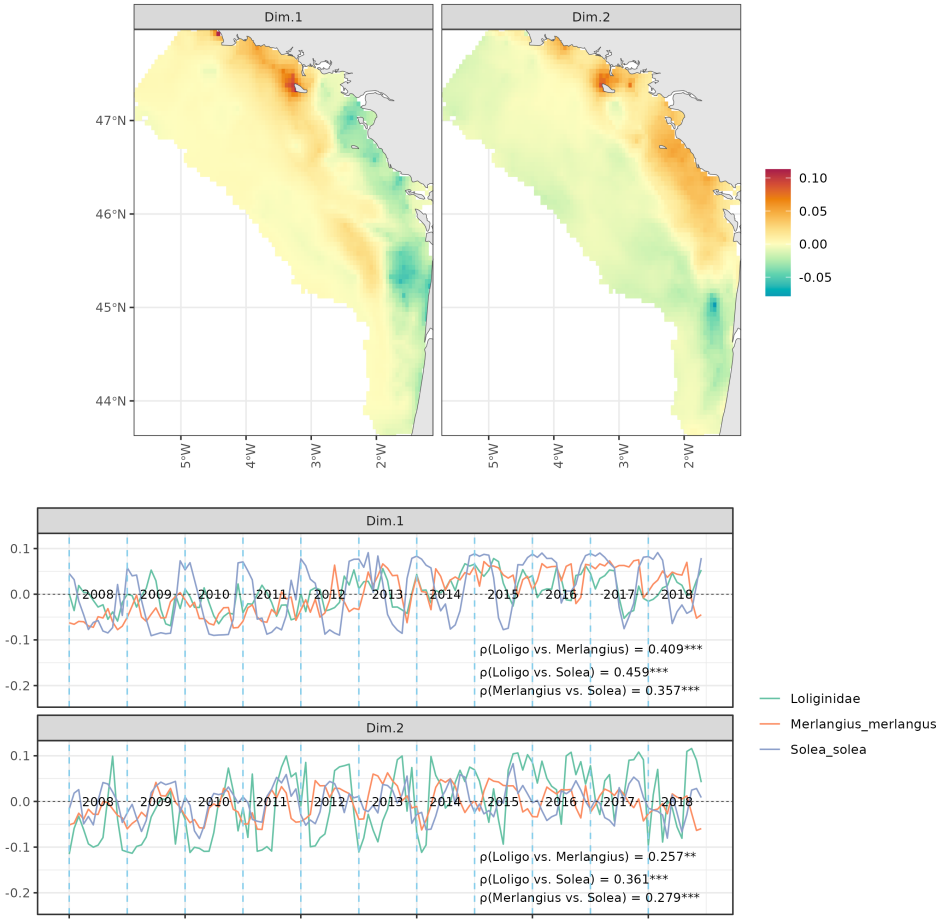

Fig. 6: IST-SDM predictions. Results for multivariate spatially synthetic EOFs (EOFs on $\mathbf{S}'^{(col)}_{multi}$). (Top) Factor maps for the first two dimensions. (Bottom) Loadings of each species for the first two dimensions. For each dimension, the loadings of the species represent when the several species of the analysis follow the related spatial factor. Blue dashed vertical lines correspond to the month of January for each year. There is a strong offshore-onshore gradient emphasized on both dimensions with coastal areas being summer grounds. All the species emphasize strong correlation relatively to the two first spatial factors.

S.1 Supplementary material

S.1.1 IST-SDM spatio-temporal predictions and anomalies

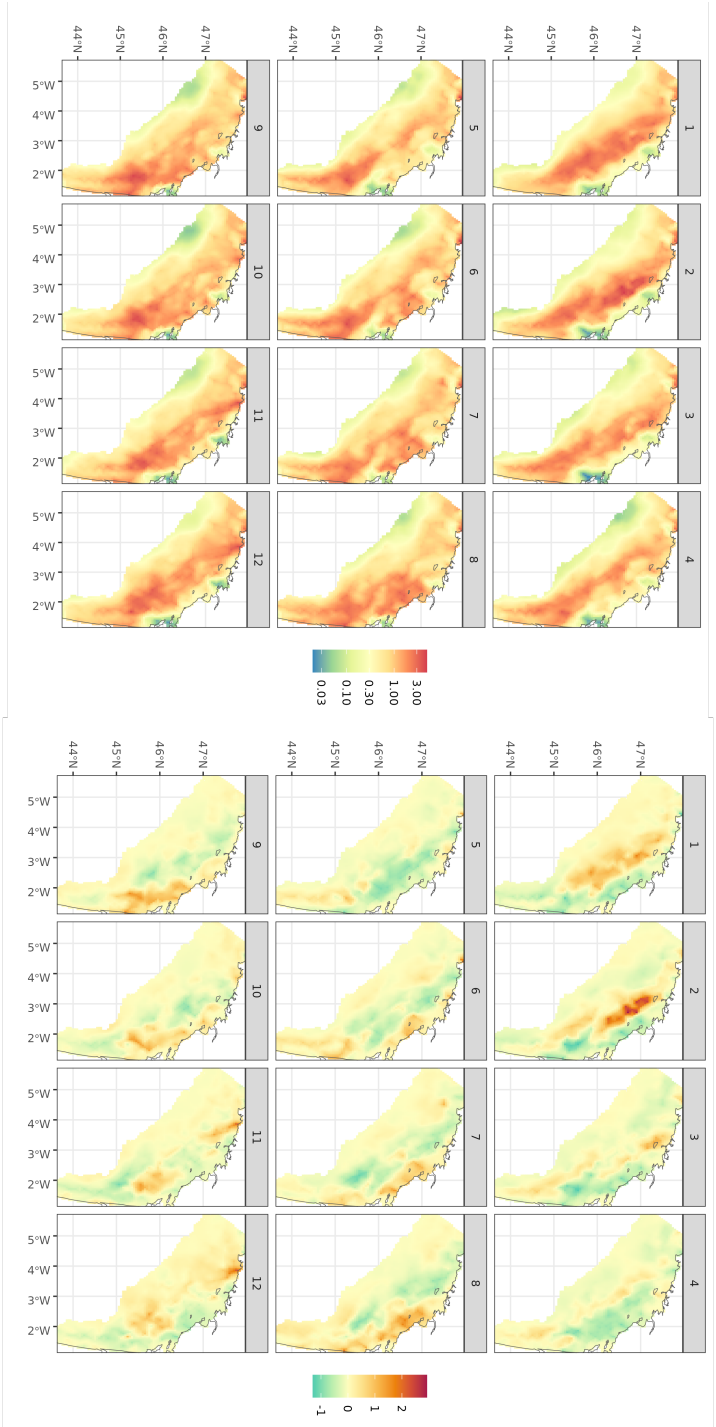

Fig. S1: IST-SDM predictions. (left) Monthly spatial predictions for sole in the log-scale $\log S(x,t)$. (right) Monthly anomalies of the spatial predictions $S^*(x,t)$. Each panel corresponds to the average anomalies for a month over the period 2008 - 2018.

S.1.2 Additionnal dimensions for EOFs

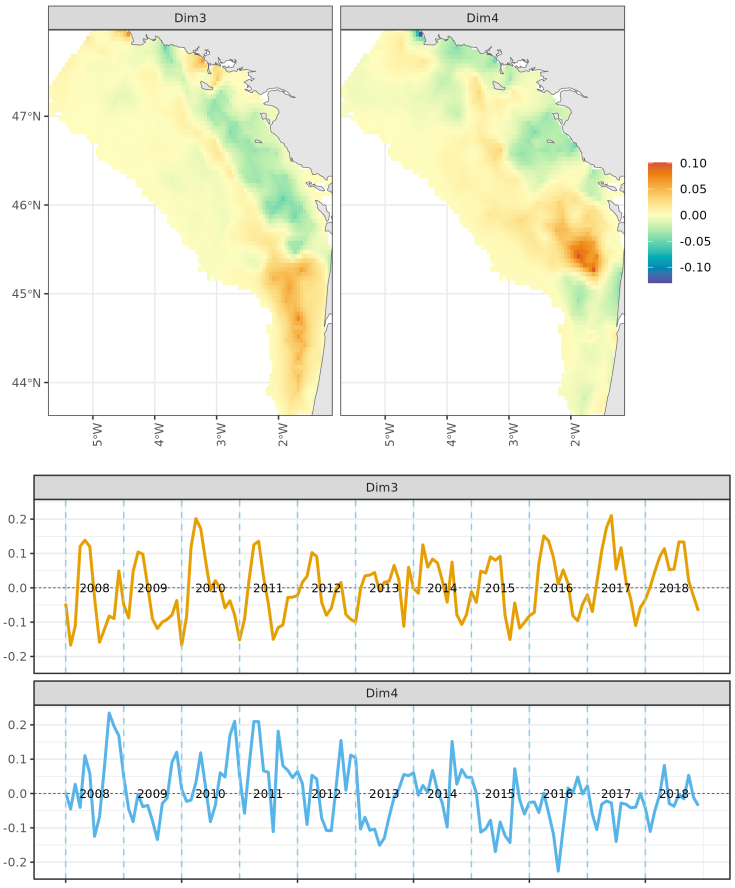

Fig. S2: IST-SDM predictions. (Top) Spatial factors for the dimensions three and four of the EOFs. (Bottom) Loadings for dimensions three and four of the EOFs. Blue dashed vertical lines correspond to the month of January for each year.

S.1.3 CCA for the IST-SDM case study

For IST-SDM predictions, we use the same parameterisation for the ancillary variable thought the delay D is fixed to $D = 0$ so that the peak of the seasonal variable falls in March (the peak of the reproduction period). This way we aim at disentangling the seasonnal

signal related to the alternance of reproduction/feeding that can be found in the four first dimensions.

The top-left panel of Figure S3 shows the resulting spatial basis vector, which reinforces the offshore-onshore gradient observed in the EOFs results (Figure 3). The inclusion of the ancillary variable enhances the clarity of these spatially distinct positive and negative regions within the canonical vector. The first three EOFs show a positive correlation with the ancillary variable, with the highest correlation coefficient being observed between CCA and the ancillary variable. This indicates that the CCA effectively extracts the temporal variation from the first three dimensions, resulting in a new variable that best aligns with the ancillary variable.

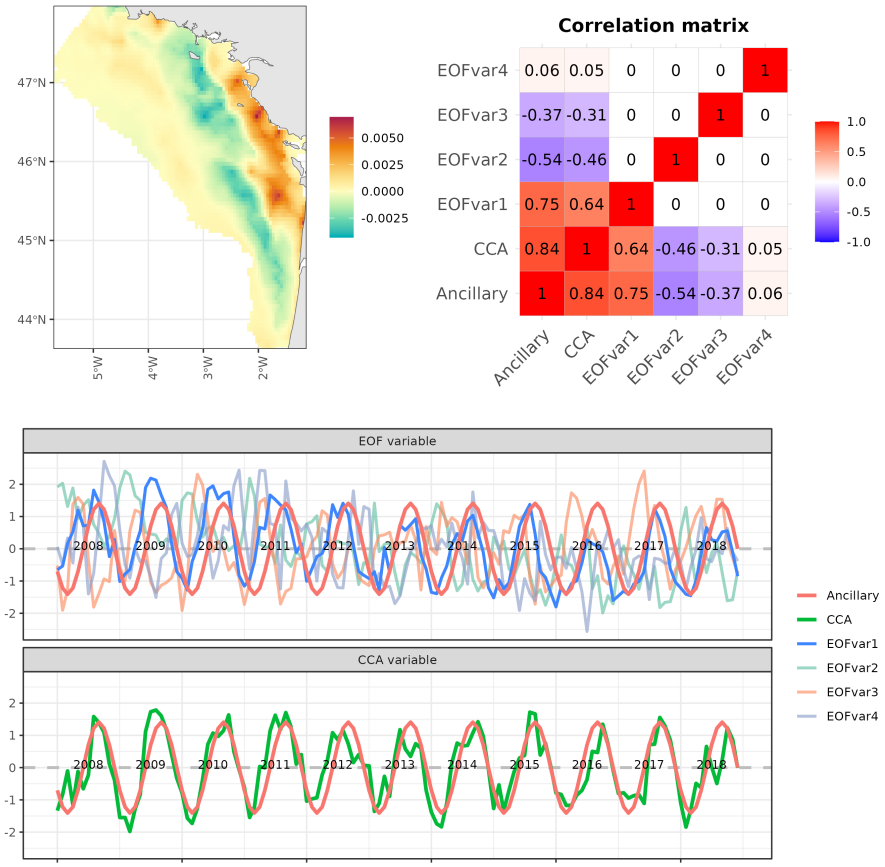

Fig. S3: IST-SDM predictions. Results of the canonical correlation analysis. (Top left) Spatial basis vectors $\mathbf{w}_1$ that maximize the correlation between the temporal variables $\mathbf{y}_1$ and $\mathbf{y}_2$. (Top right) Correlation matrix between the four first loading factors of the EOFs, the CCA $\mathbf{y}_1$, and the ancillary variable. (Bottom) Comparison of the EOFs loadings with the ancillary variable and the CCA variables with the ancillary variable. These time series are standardized. Blue dashed vertical lines correspond to the month of January for each year.

S.1.4 Lag EOFs on satellite NDVI data

S.1.4.1 Data and matrix structure

In the case of lag EOF, the matrix that is diagonalized contains the spatio-temporal data and lagged versions of the data.

Let's denote a vector of lags $\boldsymbol{\tau} = \{\tau_1, \cdots, \tau_j, \cdots, \tau_{\mathcal{T}}\}$. Then, the matrices to bind take the form:

$$\mathbf{S}'^{(\tau_j)} = \begin{pmatrix} S'(x_1, t_1 + \tau_j) & S'(x_1, t_2 + \tau_j) & \cdots & S'(x_1, t_p + \tau_j - \tau_{\mathcal{T}}) \\ S'(x_2, t_1 + \tau_j) & S'(x_2, t_2 + \tau_j) & \cdots & S'(x_2, t_p + \tau_j - \tau_{\mathcal{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ S'(x_n, t_1 + \tau_j) & S'(x_n, t_2 + \tau_j) & \cdots & S'(x_n, t_p + \tau_j - \tau_{\mathcal{T}}) \end{pmatrix}$$

with $j \in \{1, 2, \cdots, \mathcal{T}\}$

Then the more usual is to bind the $\mathcal{T}$ matrices by row (Weare and Nasstrom, 1982).

$$\mathbf{S}'_{lag} = \begin{pmatrix} \mathbf{S}'^{(\tau_1)} \\ \mathbf{S}'^{(\tau_2)} \\ \vdots \\ \mathbf{S}'^{(\tau_{\mathcal{T}})} \end{pmatrix}$$

An open question is the choice of the lags. If the data is monthly defined over several years, lags $\boldsymbol{\tau}$ can be months of the year to evidence some sequence of maps that repeat over the full time series. For instance, $\boldsymbol{\tau} = \{0, 3, 6, 9\}$ evidence sequence with quarterly lags over a

full year. $\boldsymbol{\tau} = \{0, 1, 2, \cdots, 11\}$ would evidence sequence with monthly lags over a full year. The matrix $\mathbf{S}'_{lag}$ is then diagonalized like standard EOFs.

In this case, the spatial factors ($\mathbf{u}_1$ for instance) are composed of $\mathcal{T}$ maps and the temporal loadings (*e.g.* $\mathbf{v}_1$) are vectors of length $p - \mathcal{T}$.

S.1.5 Illustration with NDVI data

We take NDVI satellite data to illustrate lag EOFs. In the example, we consider quarterly lags, so $\boldsymbol{\tau} = \{0, 3, 6, 9\}$. Based on the scree plot (Figure 2, top right), we select the three first dimensions that represents respectively 20%, 20% and 10% of the variance.

The first and second dimension have an annual period while the third one have a six-month period. This last cycle was not evidenced by the standard EOF.

First and second dimension of the lag EOFs are very similar to the one of the standard EOFs with a 3-months offset. They outline the annual variation of NDVI throughout the year with high values of NDVI in summer specifically in mountains and low values during winter.

In the third dimension, loadings are high in February/March as well as August/September and low in May/June as well as November/December (Figure S5). The areas that contribute most to the sequence of maps are areas in the North of France (Figure S6). This dimension highlights that there are peaks of vegetation in spring (April/May) and later in autumn (September/October) that can be related to the higher growth of vegetation due most likely to rain and moderate temperatures in these areas at this period of the year. By contrast, NDVI is low in July and November which are either dry periods for summer or cold period which tend to reduce the growth of vegetation.

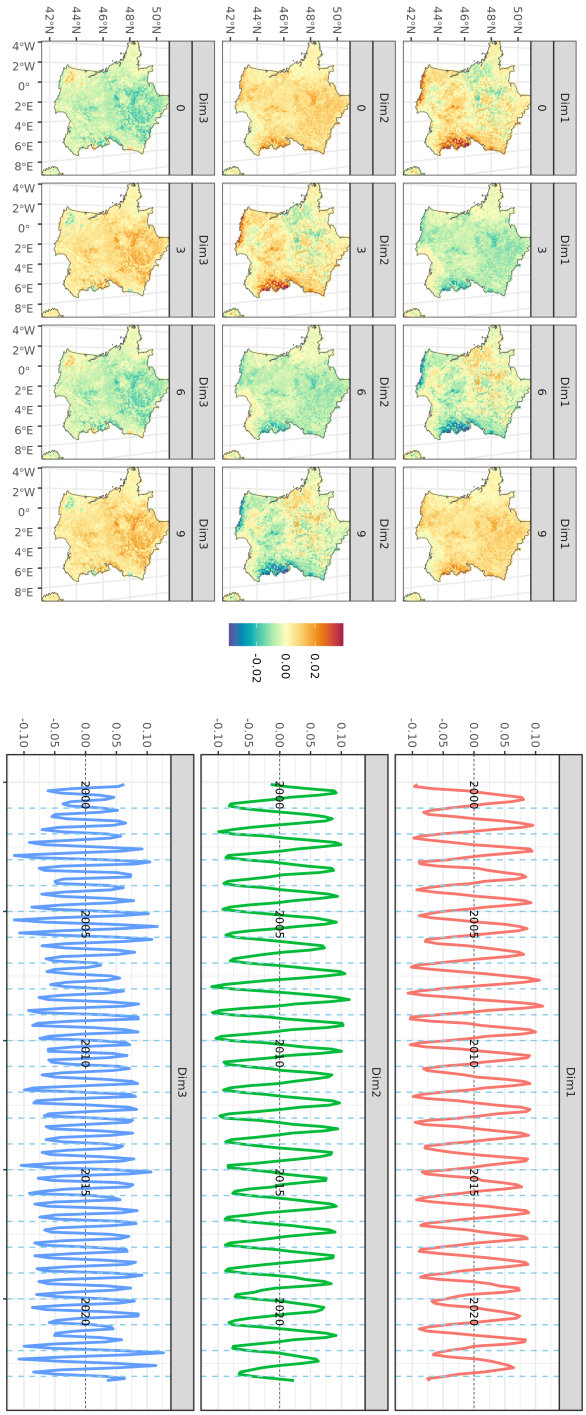

Fig. S4: Satellite NDVI data. (Top) Spatial factors for the three first dimensions of the lag EOFs. (Bottom) Loadings for the three first dimensions of the lag EOFs. Blue dashed vertical lines correspond to the month of January for each year.

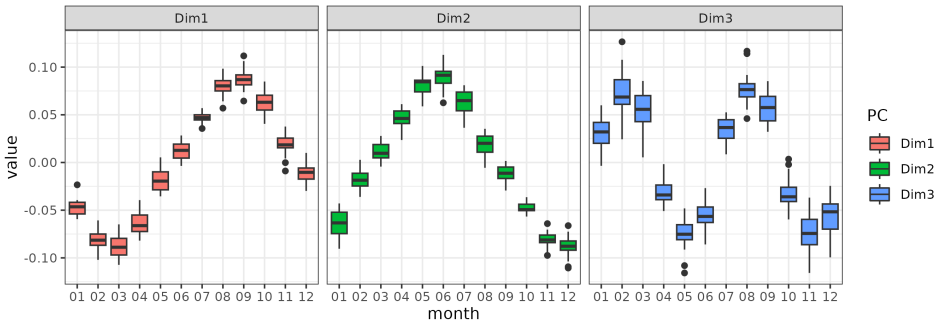

Fig. S5: Satellite NDVI data. Monthly boxplot of the temporal loadings.

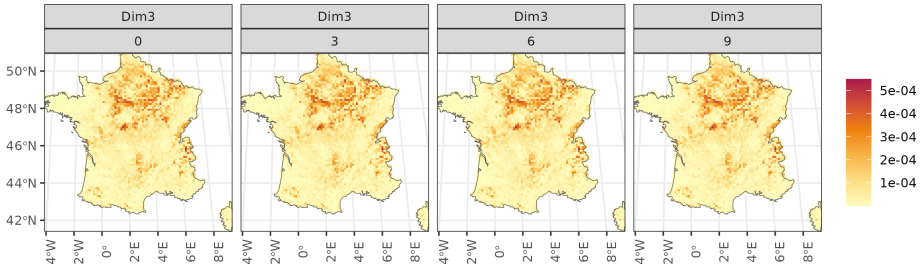

Fig. S6: Satellite NDVI data. Local variance explained by the lag EOFs for the third dimension.

S.1.6 Multispecies analysis

S.1.6.1 EOFs raw results

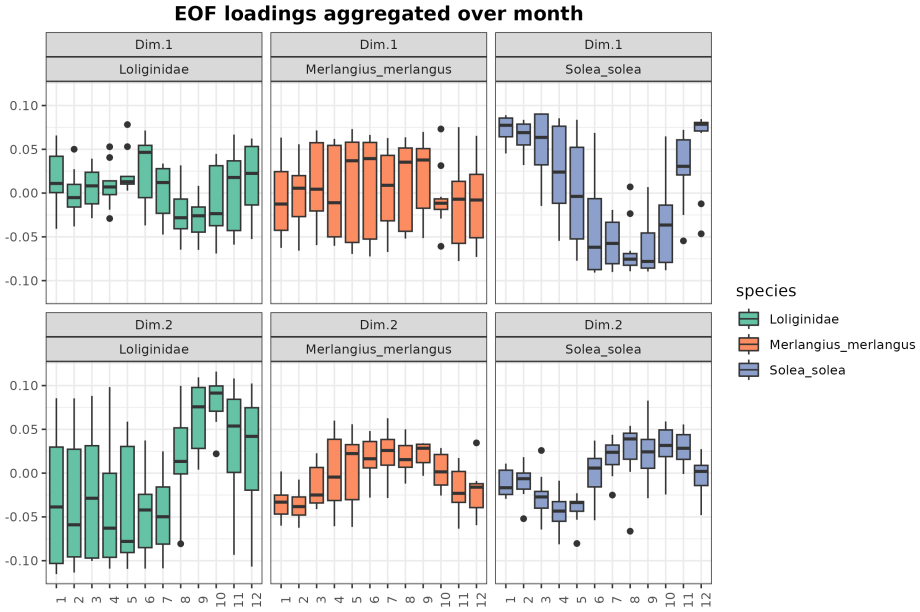

Fig. S7: IST-SDM predictions. Loadings of the multivariate analysis (Figure 6) aggregated over months.

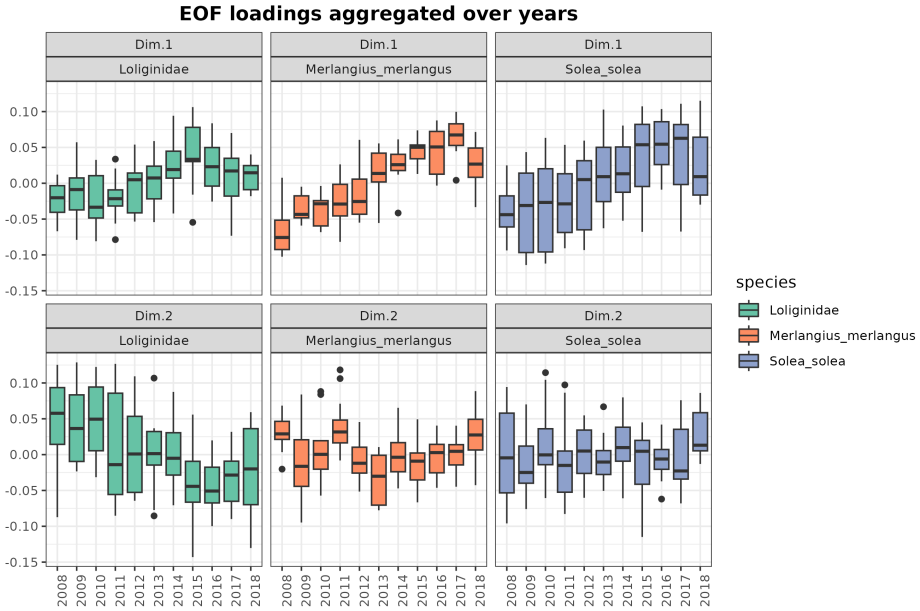

Fig. S8: IST-SDM predictions. Loadings of the multivariate analysis aggregated over years.

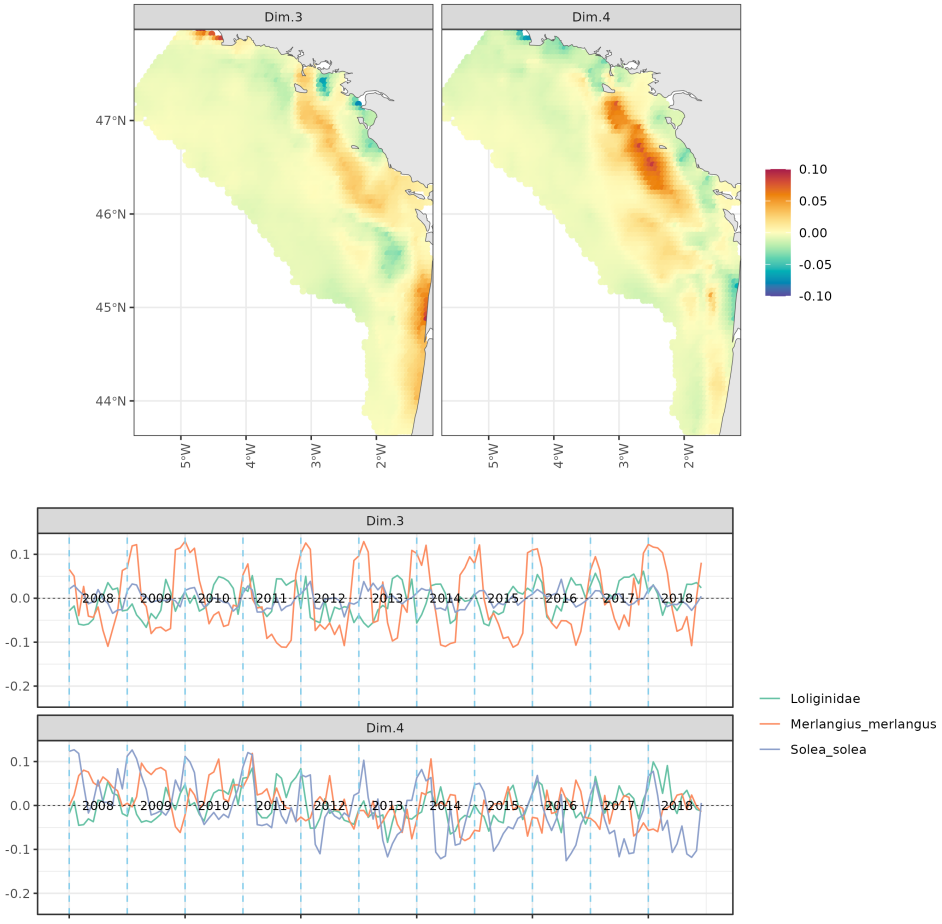

Fig. S9: IST-SDM predictions. Multivariate EOFs on the matrix $\mathbf{S}'^{(col)}_{multi}$. (Top) Spatial factors for dimensions three and four. (Bottom) Loadings of each species for dimensions three and four. For each dimension, the loadings of the species represent when the several species of the analysis follow the related spatial factor. Blue dashed vertical lines correspond to the month of January for each year.

S.1.6.2 Clustering analysis results

As mentioned in Alglave, Olmos, et al. (2024), to identify distinct essential habitats and to relate these to the ecological seasons, two complementary clustering analyses can be performed: one on the spatial factors (spatial dimension) and one on the loading factors (temporal dimension).

While the clustering in the temporal dimension identifies groups of time steps that have similar spatial patterns, the clustering in the spatial dimension identifies locations with similar fish density at the same time step. Clusters of locations can be interpreted as distinct functional habitats and clusters of time steps can be interpreted as ecological seasons. Both can be represented on a single graph and locations can be related to seasons when they are in the same direction on the graph (similarly as is done in PCA).

In a multivariate perspective, depending whether we choose the temporal synthetic representation or the spatially synthetic representation, the information that comes out can be different.

When adopting the temporally synthetic representation (EOFs on $\mathbf{S}'^{(row)}_{multi}$), a set of seasons (*i.e.* temporal clusters) is identified for all the species based on the joint temporal loadings. By contrast, the spatial clusters are built based on the spatial factors of the different species. Thus there is not only one map of spatial cluster, but as many maps of spatial clusters as there are species.

When adopting the spatially synthetic representation (EOFs on $\mathbf{S}'^{(col)}_{multi}$), a single map of spatial clusters (functional areas) is obtained. By contrast, the temporal clusters are based on the loadings of the different species.

These differences affect the ecological interpretation of the clustering. Developing both

temporally and spatially synthetic multivariate EOFs method would be relevant to provide a tighter representation of the data. We believe this goes beyond the scope of the paper and we let this for future research.

Temporally synthetic representation

The spatial clusters 3 and 5 are related to seasons B and D which are mainly winter and spring months. These are reproduction period for the species under study (Arbault, Camus, and Bec, 1986; Houise, 1993; Moreno et al., 2002); the related spatial clusters can be interpreted as reproduction areas for these species. This is consistent with the previous studies investigating the reproduction areas of these species (Alglave, Vermard, et al., 2023). Interestingly, the spatial cluster 5 is only found for whiting and highlights a specific reproduction dynamics in the North of the Bay of Biscay during winter (4.5°W - 4°N).

The spatial clusters 1 and 4 are summer grounds. They are related to seasons C and E that mainly gather late spring, summer and autumn months. The spatial cluster 1 is mainly related to late spring and summer months. The cluster 4 is related to a longer season spanning from summer to end autumn just before reproduction happens for sole and whiting.

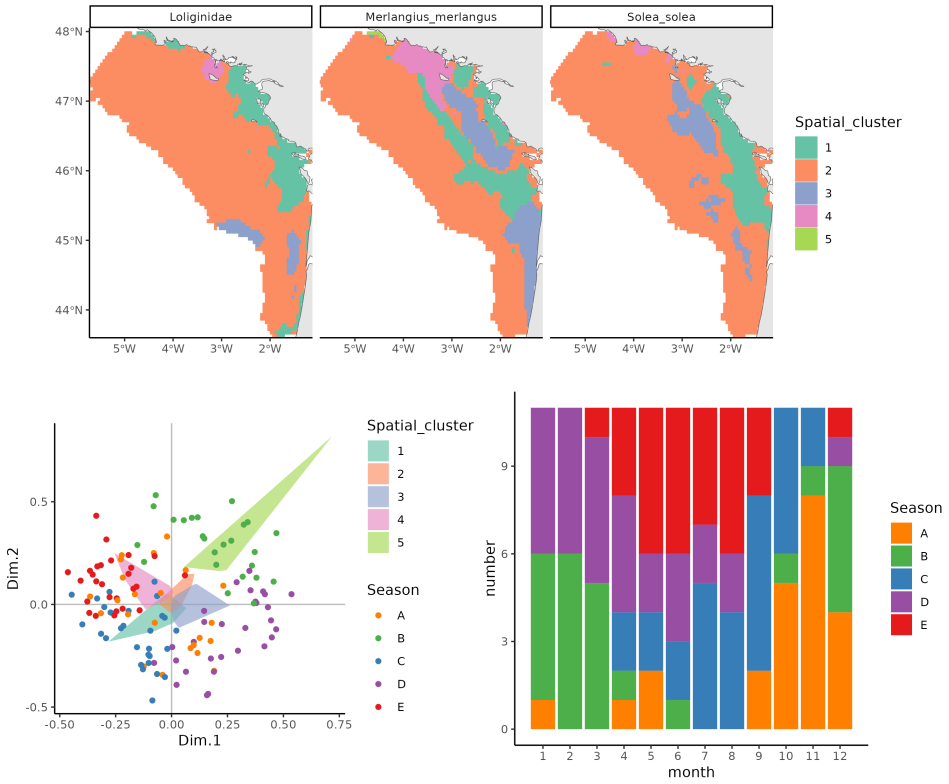

Fig. S10: Clustering analysis for the EOFs outputs of the temporally synthetic representation (EOFs on $\mathbf{S}'^{(row)}_{multi}$). (Top row) Spatial clusters representing the functional zones in space for each three species. (Bottom left) Temporal loadings and spatial clusters projected in the two first dimensions of the PCA. (Bottom right) Repartition of the different temporal clusters over the year. Each column sums to eleven *i.e.* the number of year in the analysis. Clustering trees are available below in Figure S11 (locations) and S12 (time steps).

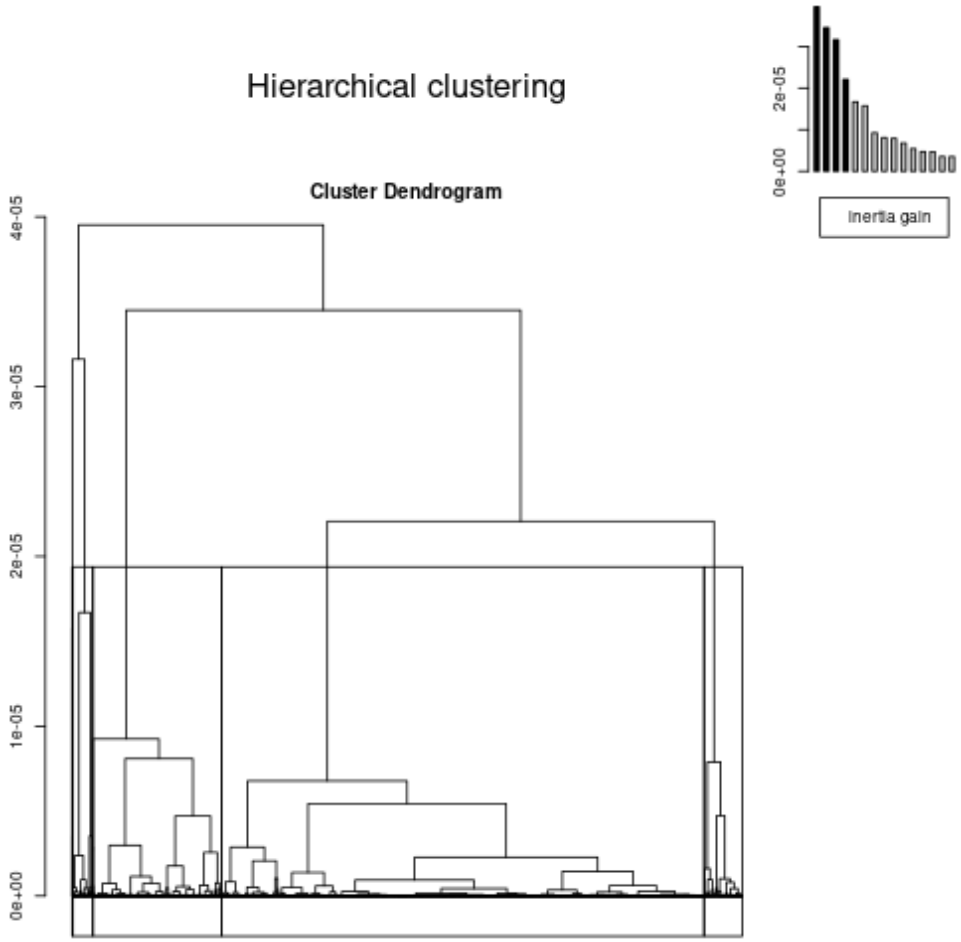

Fig. S11: Clustering tree of the spatial locations for the EOFs outputs of the temporally synthetic representation (EOFs on $\mathbf{S}'^{(row)}_{multi}$).

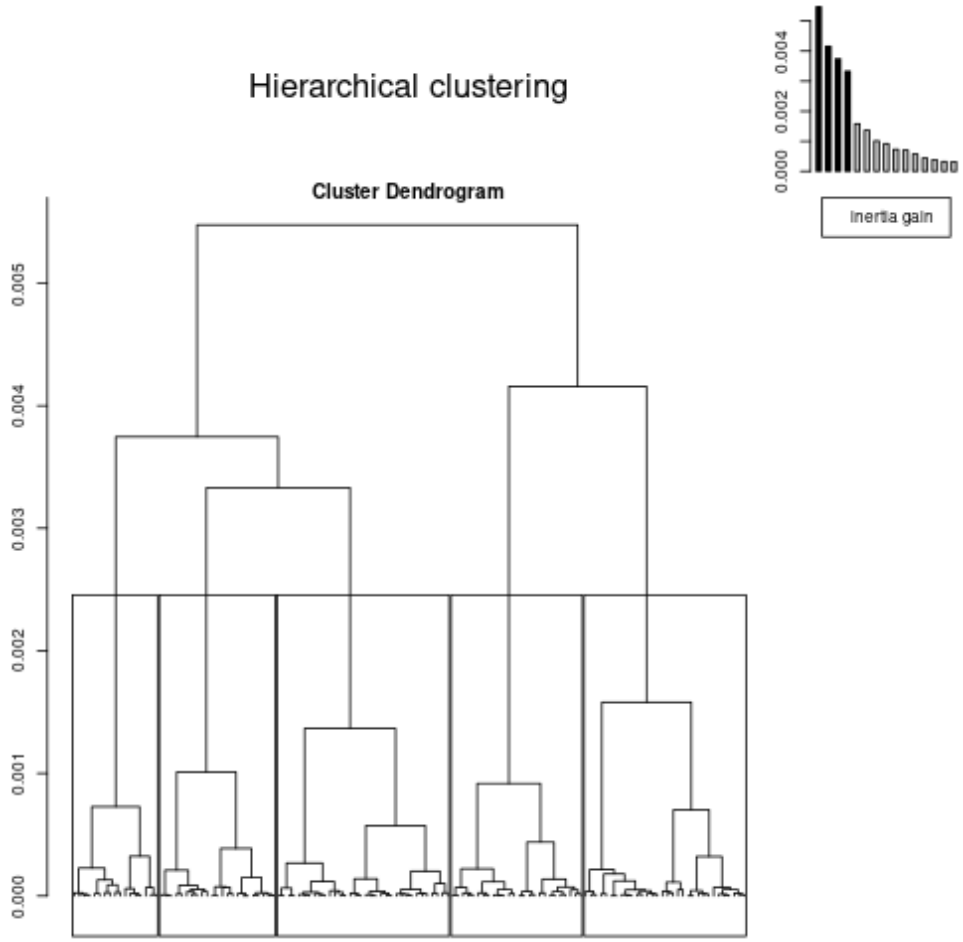

Fig. S12: Clustering tree of the temporal loadings for the EOFs outputs of the temporally synthetic representation (EOFs on $\mathbf{S}'^{(row)}_{multi}$).

Spatially synthetic representation

Five spatial clusters are identified through the spatially synthetic representation of the EOFs. The spatial cluster 1 is related to season B that fall mainly in summer and autumn for each three species. The related functional area (spatial cluster 1) can be considered as a summer grounds for each three species.

The spatial cluster 4 can be related to the time cluster A that mainly fall in late winter to early summer for squids, which includes the period of squids reproduction (January to March). For sole it falls in summer during feeding.

Interestingly, the temporal cluster D falls in winter to spring for each three species (January to June depending on the species) and coincides with the reproduction period of the different species. For whiting, the cluster E too falls during reproduction. For sole, this coincides with the migration off the coast and to the North *i.e.* off spatial cluster 1 and 4 towards spatial clusters 2 and 5.

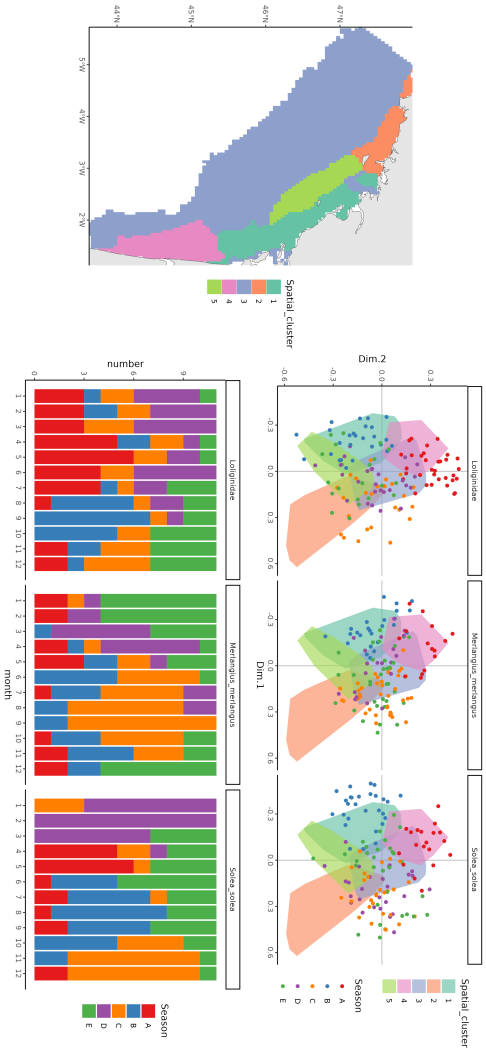

Fig. S13: Clustering analysis for the EOFs outputs of the spatially synthetic representation (EOFs on $\mathbf{S}'^{(col)}_{multi}$). (Left) Spatial clusters representing the functional zones in space. (Top row) Projection of the temporal loadings and spatial locations in the two first dimensions of the PCA. The clusters of the locations are the distinct functional zones and the clusters of the temporal loadings can be seen as ecological seasons. (Bottom row) Repartition of the different temporal clusters over the year. Each column sums to eleven *i.e.* the number of year in the analysis. Clustering trees are available below in Figure S14 (locations) and S15 (time steps).

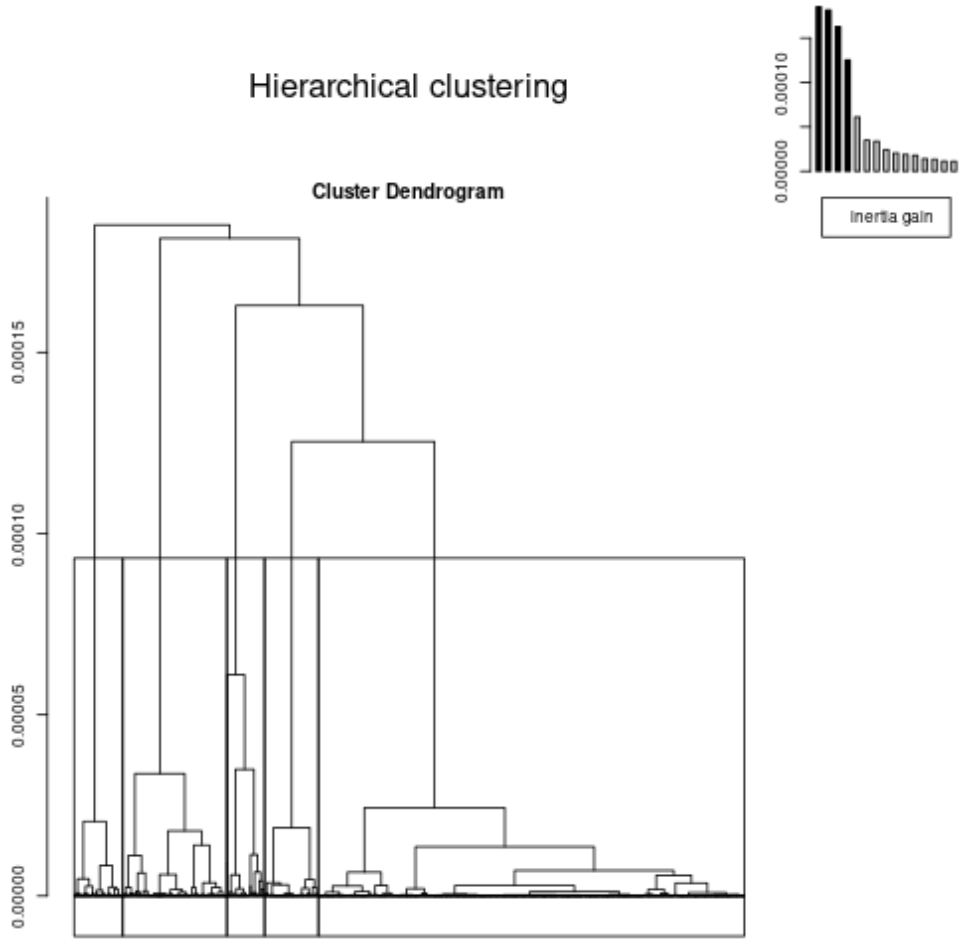

Fig. S14: Clustering tree of the spatial locations for the EOFs outputs of the spatially synthetic representation (EOFs on $\mathbf{S}'^{(col)}_{multi}$).

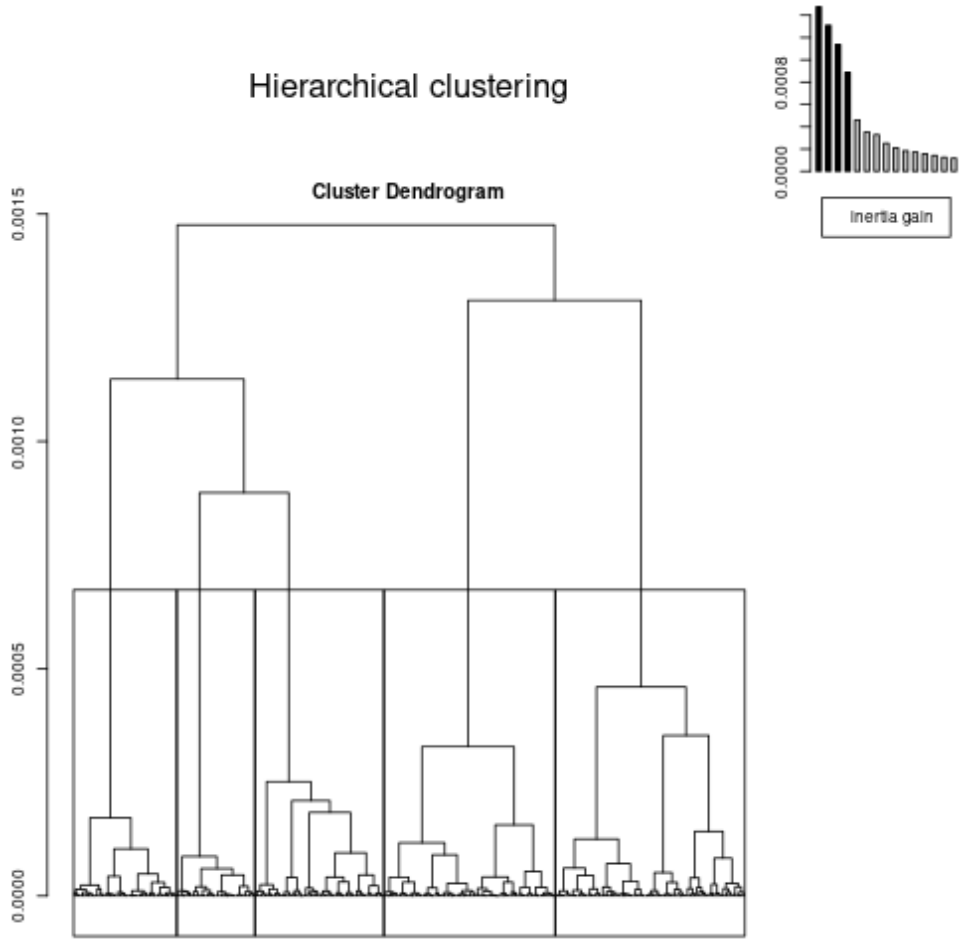

Fig. S15: Clustering tree of the temporal loadings for the EOFs outputs of the spatially synthetic representation (EOFs on $\mathbf{S}'^{(col)}_{multi}$).