



**HAL**  
open science

## Empirical Orthogonal Functions for ecology

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

► **To cite this version:**

Baptiste Alglave, Benjamin Dufée, Said Obakrim, James T Thorson. Empirical Orthogonal Functions for ecology. 2024. hal-04693871

**HAL Id: hal-04693871**

**<https://hal.science/hal-04693871v2>**

Preprint submitted on 16 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1 **Empirical Orthogonal Functions for ecology**

2 Baptiste Alglave<sup>1,2</sup>, Benjamin Dufée<sup>1,3</sup>, Said Obakrim<sup>4</sup>, and James T. Thorson<sup>5</sup>

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

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

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

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

23 Finally, EOFs can be a valuable tool for investigating community and ecosystem spatio-temporal  
24 dynamics, though this implies to move to a multivariate framework. We illustrate several ways to  
25 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,

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

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

## 32 **1 Introduction**

33 Spatial and temporal data are ubiquitous in ecological research, providing insights across  
34 various scales. Survey data, analyzed through multivariate integrated modeling, help  
35 understanding species distribution, ecosystem states and community dynamics (Isaac et al.,  
36 2020; Zipkin et al., 2021). Citizen science initiatives yield extensive opportunistic datasets  
37 over long periods, facilitating studies on phenomena such as bird migration (Martin et al.,  
38 2020; Sullivan et al., 2014). GPS data allows to track animal movements at high resolution  
39 (Nathan et al., 2022). Remote sensing offers high-resolution information for monitoring  
40 land use, habitat suitability, and the impacts of climate change in near real-time (Poggi et al.,  
41 2021; Garrigues, Allard, and Baret, 2008). Together, these diverse spatio-temporal data  
42 sources are crucial for unraveling spatial ecological processes, their temporal trends, and  
43 the interactions affecting ecological processes across various scales.

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

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

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

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

68 Researchers have also applied EOFs to species' spatio-temporal distributions (with a specific  
69 emphasis on fish and marine ecosystems), linking them to climate variables, particularly in  
70 studies of sardines and anchovies in the California Current system (Norton and Mason, 2005)  
71 or in the Bay of Biscay (Petitgas, Masse, et al., 2006; Petitgas, 2008; Petitgas, Doray, et al.,  
72 2014). Some research used MFA (a derived method of PCA) to explore the spatial structure  
73 of the ecosystem based on multi-trophic observations (Woillez et al., 2010; Petitgas, Huret,

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

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

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

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

92 EOFs do not incorporate any ecological constraints, and their results may not always have a  
93 direct ecological interpretation due to standard PCA constraints (*i.e.* orthogonality). Using  
94 canonical correlation analysis (CCA), we impose an ecological constraint on EOFs to  
95 improve their ecological interpretability. Additionally, EOFs have proven to be valuable  
96 tools for community modeling and for ecosystem dynamics analysis (Petitgas, Huret, et al.,  
97 2018; Thorson, Arimitsu, et al., 2021). Here, we introduce two approaches to perform

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

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

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

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

## 124 **2 Basics of EOFs**

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

### 130 **2.1 The raw spatio-temporal field**

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

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

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

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

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

136 Examples of both the raw spatio-temporal data and anomalies are available for satellite  
 137 NDVI data and IST-SDM predictions in Figure 1 and S1. Codes below are provided to  
 138 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)
```

139



## 140 2.2 EOFs formalism: decomposing the spatio-temporal field

### 141 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:

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

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

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

153 **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:

$$\mathbf{S}' = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \quad (4)$$

154 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  
155  $\mathbf{S}'$ .  $\mathbf{\Lambda}$  is a  $r \times r$  diagonal matrix with non-increasing positive coefficients on the diagonal,  
156 denoted as  $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_r)$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ . This is essentially the same  
157 computation that is performed when conducting PCA on any type of multivariate data.

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

161 The columns of  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_r)$  are often referred to as loading factors in PCA terms.  
162 Here, they will be called either loadings or temporal loadings of the anomaly matrix  $\mathbf{S}'$ .  
163 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:

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

164 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:

$$I_m = \lambda_m^2$$

165 The matrix decomposition of equation (4) coincides with equation (3) provided the spatial  
166 factors  $\mathbf{u}_m$  are normalized by the singular values. Below are the codes to perform SVD in R  
167 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)
```

168

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

### 176 2.2.3 Interpretation of EOFs

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

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

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

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

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

## 222 **3 Data and case study for illustration**

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

### 227 **3.1 Satellite NDVI data**

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

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

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

### 248 **3.2 Species distribution model predictions**

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

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

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

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

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

#### 272 **4 Applying EOFs to the case studies**

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

279 For satellite NDVI data, both the first and second dimensions emphasize seasonality on  
280 two different phases. The first dimension is related to the winter/summer seasonality with  
281 a positive peak in June and a progressive decrease until February/March. This signal is  
282 mainly related to the mountainous and eastern part of France. It is basically the signal that  
283 represents (1) the decrease of vegetation in winter (especially in mountainous areas) due  
284 to smaller primary production and snow falls and (2) an increase starting in March when  
285 temperatures are higher and conditions are favorable for the growth of vegetation.

286 By contrast, the second dimension emphasizes another phase in seasonality. Temporal  
287 loadings are negative from March to June (spring and early summer), which corresponds  
288 to high NDVI in the western and northern part of France. This can be related to both rain  
289 and higher temperature conducting to high primary production in this area. From July to



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

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

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

## 308 **5 Informing EOFs with an ancillary temporal variable**

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

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

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

## 322 **5.1 Basics of Canonical correlation analysis**

323 CCA was first proposed by Hotelling (1992). It is a method used to identify and measure the  
324 associations among two sets of variables. Consider two space-time processes  $S^{(1)}(x, t)$  and  
325  $S^{(2)}(x, t)$  at location  $x$  and time  $t$ . CCA aims at identifying pairs of spatial basis vectors,  $\mathbf{w}_1$   
326 and  $\mathbf{w}_2$ , for the two processes,  $S^{(1)}$  and  $S^{(2)}$ , that maximize the correlation between their  
327 projections onto these vectors. Consider the corresponding observed space-time matrices  
328 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  $\rho$  between  $\mathbf{y}_1 = \mathbf{S}^{(1)}\mathbf{w}_1$  and  $\mathbf{y}_2 = \mathbf{S}^{(2)}\mathbf{w}_2$ :

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

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

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

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

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

## 335 **5.2 Coupling CCA with EOFs**

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

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

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

### 349 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:

$$An(t) = A \cdot \sin(f \cdot t + D) \quad (8)$$

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

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

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

### Perform CCA between EOFs results and a seasonal time series

```
## Construct the seasonal 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]
```

358

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

360 = 0.98, Figure 4, top right). By comparison, the first two EOFs loadings have positive, but  
361 lower correlation coefficient ( $R^2 = 0.82$  and  $R^2 = 0.57$ ). The spatial basis obtained through  
362 CCA displays a stronger North/South gradient of NDVI compared with the two first EOFs  
363 (Figure 3). NDVI are high in orange/red areas during summer (i.e. the center and mountain  
364 part of France) which corresponds to high NDVI in mountains and all the southern part and  
365 eastern part of France. In these areas, either ice melt and temperatures become favorable for  
366 vegetation growth (mountains), either temperature are high and rains remain frequent which  
367 increases vegetation coverage (eastern and south western part of France). In winter, these  
368 areas are colder and they can be covered by snow which do not favor vegetation growth.

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

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

## 374 **6 EOFs for multivariate analysis**

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

382 **6.1 The two approaches for conducting multivariate EOFs**

383 Let  $v \in \{1, \dots, s\}$  denotes the index representing the number of variables in the analysis,  
 384 and let  $S^{(v)}(x, t)$  denotes the value of the space-time process for location  $x$ , time  $t$ , and  
 385 variable  $v$ . We denote by  $\mathbf{S}^{(v)}$  the  $n \times p$  matrix of the spatio-temporal variable  $v$ .

386 Two options are available when conducting multivariate EOFs:

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

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

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

$$\mathbf{S}_{multi}^{(col)} = \left( \mathbf{S}^{(1)} \ ; \ \mathbf{S}^{(2)} \ ; \ \dots \ ; \ \mathbf{S}^{(s)} \right)$$

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

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



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

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

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

## 412 **6.2 Illustration**

413 To illustrate multivariate EOFs analysis, we apply the two alternative approaches to the  
414 three species in the Bay of Biscay: sole, whiting, and squids. These are coastal species  
415 (Alglave, Vermard, et al., 2023) with potentially shared spatio-temporal dynamics *e.g.*  
416 same reproduction period (Arbault, Camus, and Bec, 1986; Houise, 1993; Moreno et al.,  
417 2002). From an ecological perspective, a critical issue involves identifying shared spatial or  
418 temporal dynamics among different species (Thorson, Scheuerell, et al., 2015; Ovaskainen  
419 and Abrego, 2020; Thorson, Arimitsu, et al., 2021). Additionally, one can seek to identify  
420 combinations of 'areas x seasons' that can be interpreted as multispecies functional habitats  
421 (Delage and Le Pape, 2016). To do this, we combine the EOFs (both representations)

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

426 The temporally synthetic multivariate EOFs (EOFs on  $\mathbf{S}_{multi}^{(row)}$ ) provides seasonal temporal  
427 loadings that represent the reproduction phenology for all the species (as mentioned in  
428 Section 3). We select the first two dimensions (Figure 2). For sole, the reproduction areas  
429 are generally consistent with those identified when conducting univariate EOFs (Figure 5).  
430 For squids and whittings, reproduction areas are observed along the Landes coast (45°N -  
431 1.5°W), on the Vendée coast (46.5°N - 2°W), and in the north of the Bay of Biscay (47.5°N  
432 - 3°W). The clustering confirms these results (Figure S10) with an additional potential  
433 reproduction area off the continental shelf on the slope (3°W - 45°N).

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

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

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

## 449 **7 Discussion**

### 450 **Towards a wider use of EOFs in ecology**

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

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

### 460 **Accounting for spatio-temporal correlation in EOFs**

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

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

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

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

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

#### 486 **Informing EOFs with ancillary variables for a better ecological interpretation**

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

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

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

#### 509 **EOFs for community and ecosystem analysis**

510 The use of EOFs for multivariate data has gained significant importance over the past  
511 decade. Initially, research focused on synthesizing information from temperature-salinity  
512 diagrams to reduce the complexity of data representation (Fukumori and Wunsch, 1991;  
513 Sparnocchia, Pinardi, and Demirov, 2003). More recently, EOFs have been applied to study  
514 ecosystem structure and community dynamics (Wuillez et al., 2010; Petitgas, Huret, et al.,  
515 2018; Grandremy et al., 2023) and to forecast these communities and ecosystems (Thorson,  
516 Arimitsu, et al., 2021).

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

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

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

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

563 **References**

- 564 Alglave, Baptiste, Maxime Olmos, et al. (2024). “Investigating fish reproduction phenol-  
565 ogy and essential habitats by identifying the main spatio-temporal patterns of fish  
566 distribution”. In: ICES Journal of Marine Science, fsae099.
- 567 Alglave, Baptiste, Etienne Rivot, et al. (2022). “Combining scientific survey and commercial  
568 catch data to map fish distribution”. In: ICES Journal of Marine Science 79.4, pp. 1133–  
569 1149.
- 570 Alglave, Baptiste, Youen Vermard, et al. (2023). “Identifying mature fish aggregation areas  
571 during spawning season by combining catch declarations and scientific survey data”. In:  
572 Canadian Journal of Fisheries and Aquatic Sciences 80.5, pp. 808–824.
- 573 Arbault, Par Suzanne, P Camus, and C le Bec (1986). “Estimation du stock de sole (*Solea*  
574 *vulgaris*, Quensel 1806) dans le Golfe de Gascogne à partir de la production d’oeufs”.  
575 In: Journal of applied ichthyology 2.4, pp. 145–156.
- 576 Astigarraga, Julen et al. (2020). “Evidence of non-stationary relationships between climate  
577 and forest responses: Increased sensitivity to climate change in Iberian forests”. In:  
578 Global Change Biology 26.9, pp. 5063–5076.
- 579 Berkelhammer, Max (2019). “Synchronous modes of terrestrial and marine productivity in  
580 the North Pacific”. In: Frontiers in Earth Science 7, p. 73.
- 581 Bez, Nicolas, Didier Renard, and Dedah Ahmed-Babou (2023). “Empirical Orthogonal  
582 Maps (EOM) and Principal Spatial Patterns: Illustration for Octopus Distribution Off  
583 Mauritania Over the Period 1987–2017”. In: Mathematical Geosciences 55.1, pp. 113–  
584 128.
- 585 Bi, Xuan et al. (2021). “Tensors in statistics”. In: Annual review of statistics and its application  
586 8.1, pp. 345–368.



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

590 Bueso, Diego, Maria Piles, and Gustau Camps-Valls (2020). “Nonlinear PCA for spatio-  
591 temporal analysis of Earth observation data”. In: IEEE Transactions on Geoscience and Remote Sensing  
592 58.8, pp. 5752–5763.

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

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

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

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

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

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

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

610 Eastman, J Ronald et al. (2013). “Global trends in seasonality of normalized difference  
611 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.

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

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

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

644 Iida, Takahiro and Sei-Ichi Saitoh (2007). “Temporal and spatial variability of chlorophyll  
645 concentrations in the Bering Sea using empirical orthogonal function (EOF) analysis of  
646 remote sensing data”. In: Deep Sea Research Part II: Topical Studies in Oceanography  
647 54.23-26, pp. 2657–2671.

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

650 Le Mezo, Priscilla et al. (2016). “Natural variability of marine ecosystems inferred from a  
651 coupled climate to ecosystem simulation”. In: Journal of Marine Systems 153, pp. 55–  
652 66.

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

655 — (2012b). Numerical ecology. Vol. 24. Elsevier.

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

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

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

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

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

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

669 Mestas-Nuñez, Alberto M (2000). “Orthogonality properties of rotated empirical modes”. In:  
670 International Journal of Climatology: A Journal of the Royal Meteorological Society  
671 20.12, pp. 1509–1516.

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

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

678 Moreno, A et al. (2002). “Biological variation of *Loligo vulgaris* (Cephalopoda: Loliginidae)  
679 in the eastern Atlantic and Mediterranean”. In: Bulletin of Marine Science 71.1, pp. 515–  
680 534.

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

683 Norton, JERROLD G and JANET E Mason (2005). “Relationship of California sardine  
684 (*Sardinops sagax*) abundance to climate-scale ecological changes in the California  
685 Current system”. In: California Cooperative Oceanic Fisheries Investigations Report  
686 46, p. 83.

687 Obukhov, Alexander M (1947). “Statistically homogeneous fields on a sphere”. In:  
688 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, Loic 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.

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



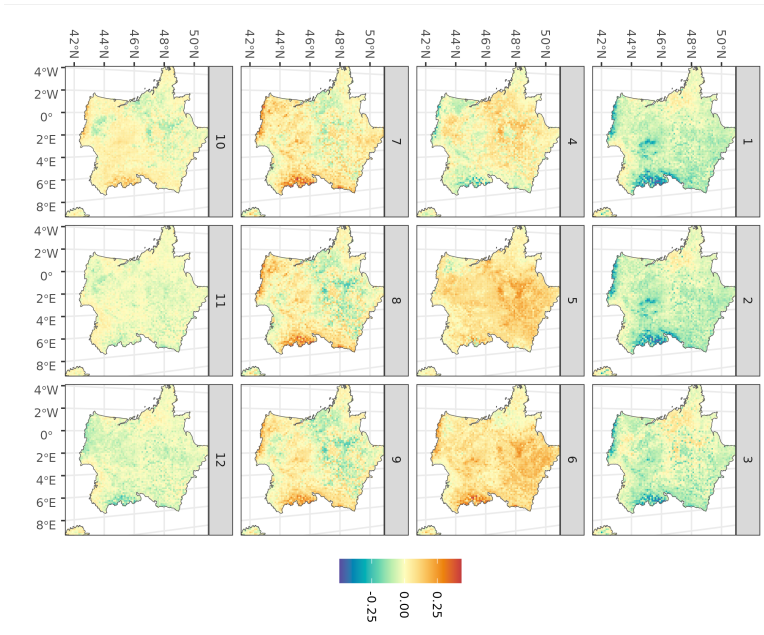
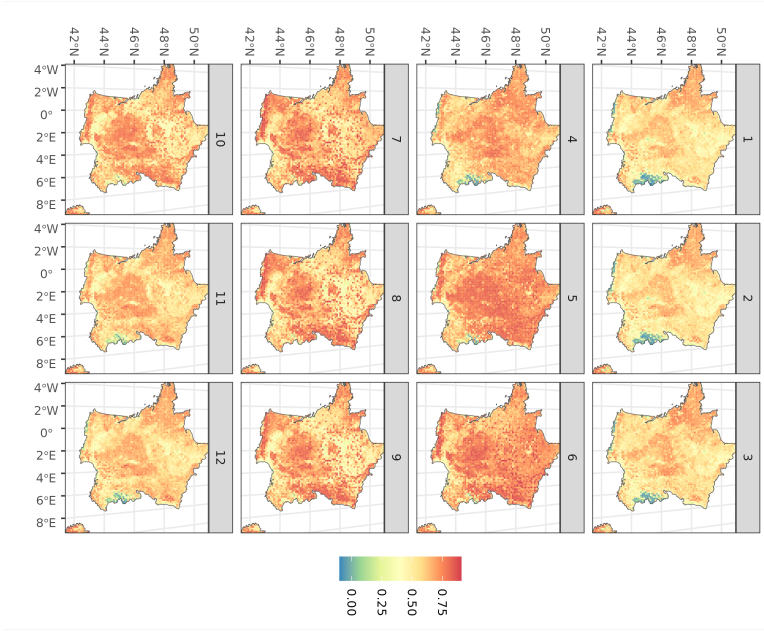


Fig. 1: NDVI data. (left) Monthly spatial data of NDVI  $S(x, t)$ . (right) Monthly anomalies of the spatial predictions  $S^*(x, t)$ . Each panel corresponds to the average predictions or anomalies for a month over the period 2000 - 2023.

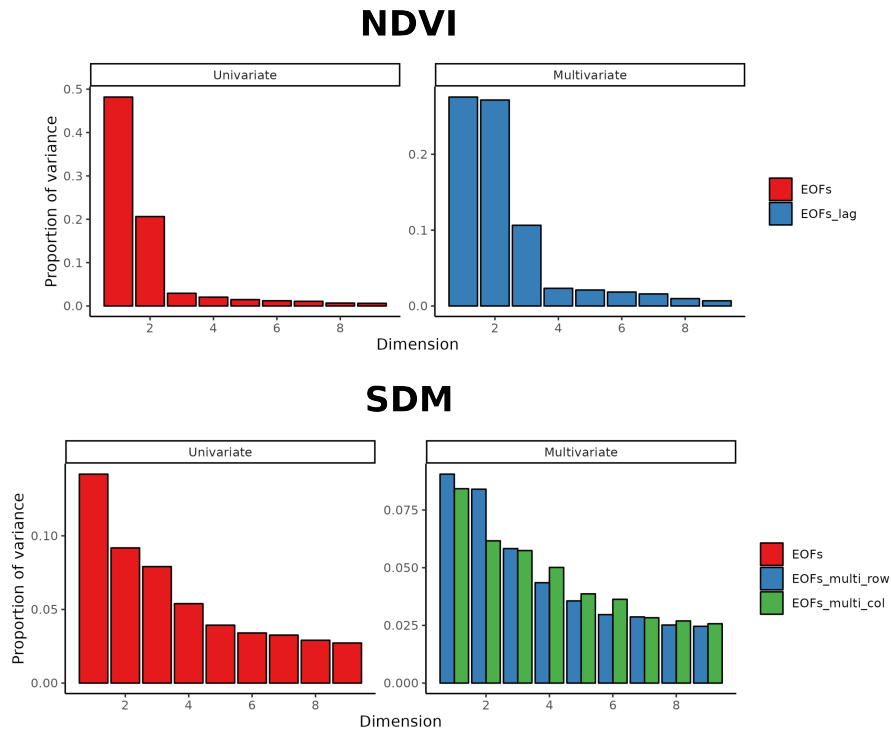


Fig. 2: Proportion of variance explained by each method for each case study. Top: NDVI satellite data, bottom: IST-SDM predictions for sole. (Left) Univariate analysis. (Right) Multivariate analysis. EoFs\_lag is the lagged version of EoFs. EoFs\_multi\_row is the multivariate EoFs conducted on  $S_{multi}^{(row)}$  (temporally synthetic representation). EoFs\_multi\_col is the multivariate EoFs conducted on  $S_{multi}^{(col)}$  (spatially synthetic representation).

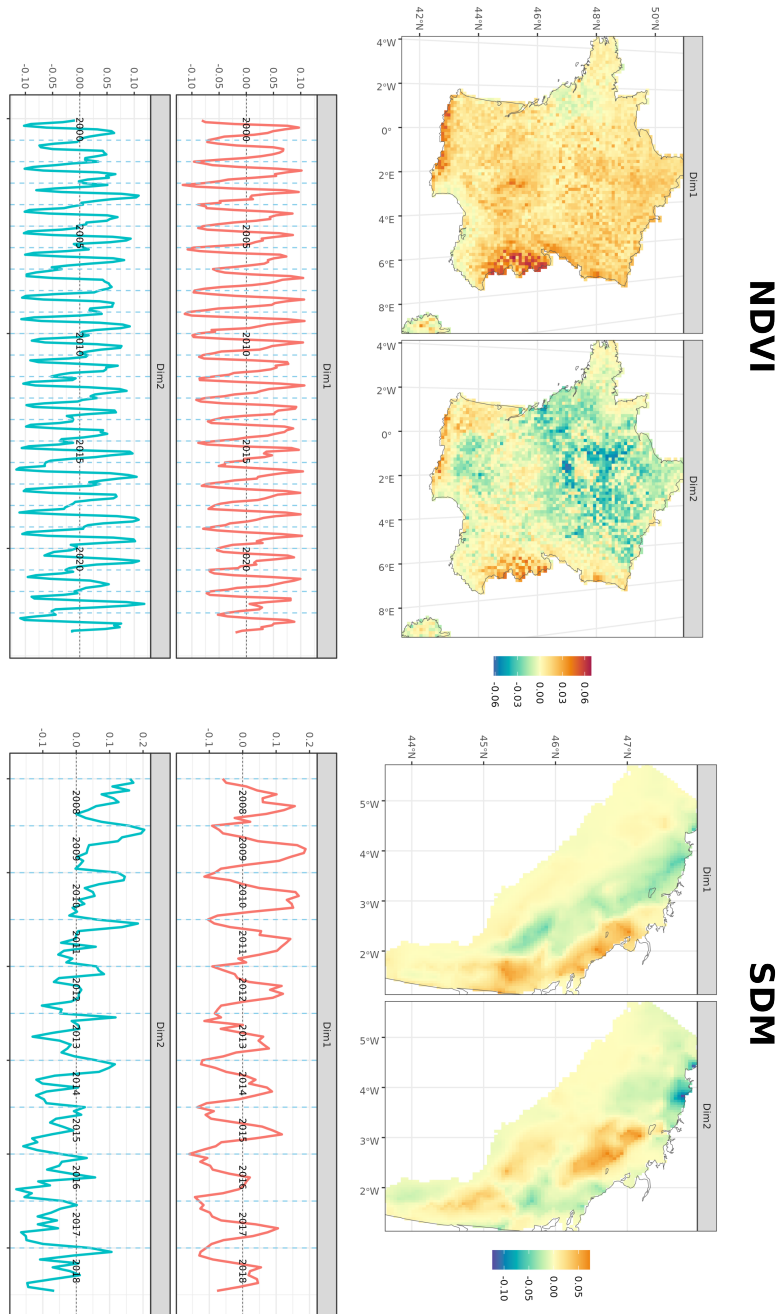


Fig. 3: EOFs results for satellite NDVI data (left) and IST-SDM predictions (right). (Top) Spatial factors for the two first dimensions of the EOFs. (Bottom) Loadings for the two first dimensions of the EOFs. Blue dashed vertical lines correspond to the month of January for each year.

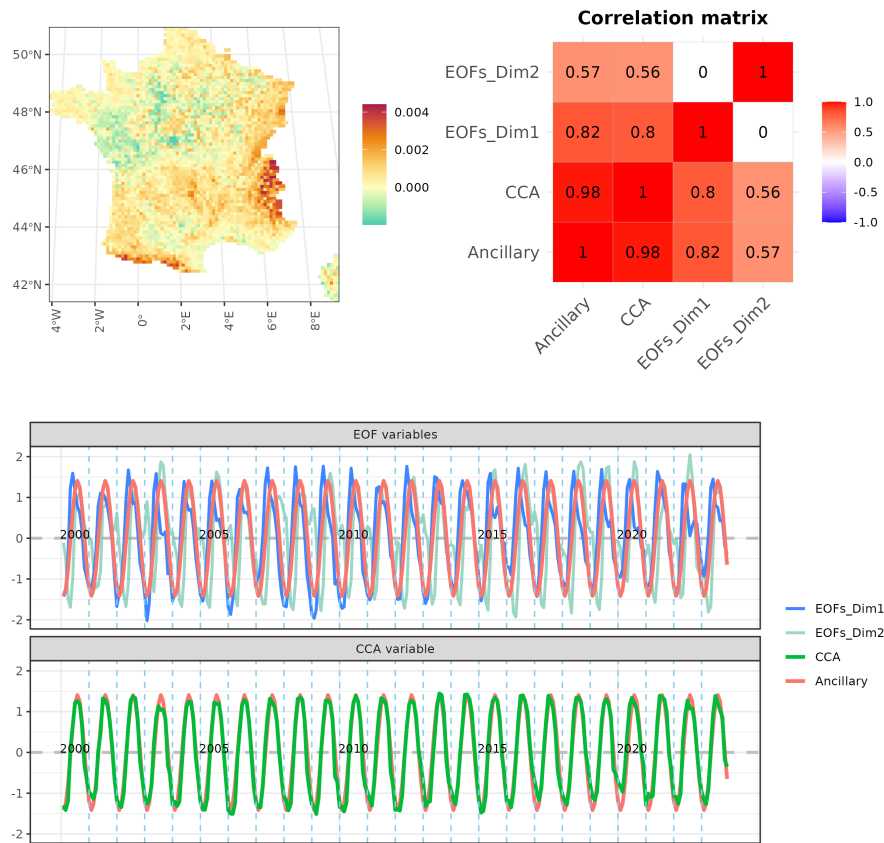


Fig. 4: Satellite NDVI data. Results of the canonical correlation analysis. (Top left) Spatial basis vector that maximize the correlation between the temporal variables  $y_1$  and  $y_2$ . (Top right) Correlation matrix between the two first loading factors of the EOFs, the CCA  $y_1$ , and the ancillary 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.

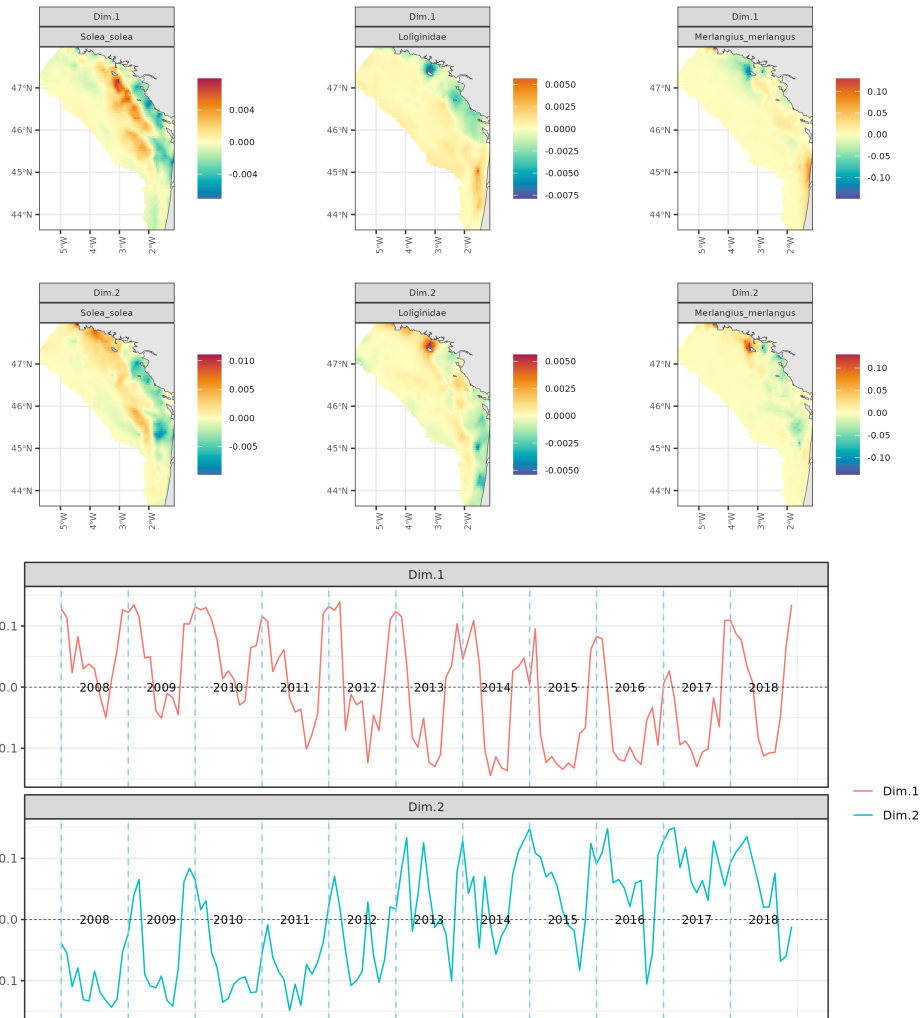


Fig. 5: IST-SDM predictions. Results for the multivariate temporally synthetic EOFs (EOFs on  $\mathbf{S}_{multi}^{(row)}$ ). (Top) Factor maps for each species and first two dimensions. (Bottom) Loadings for the first two dimensions. For each dimension, loadings represent when the several species of 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.

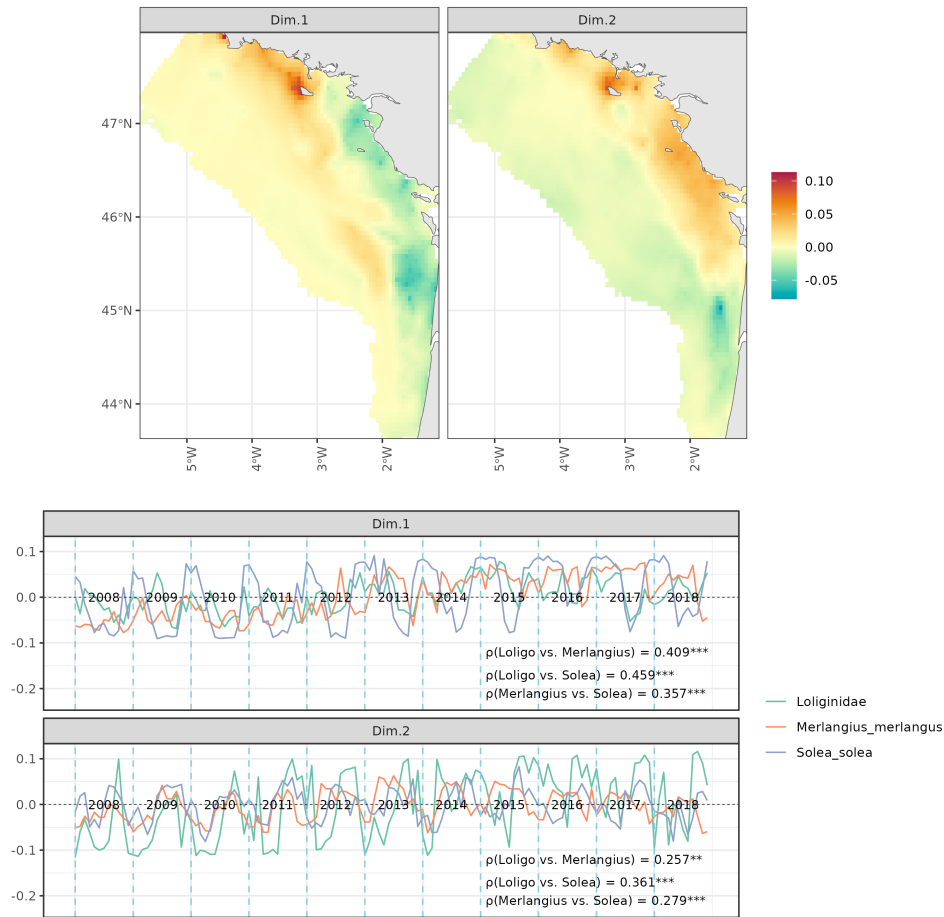


Fig. 6: IST-SDM predictions. Results for multivariate spatially synthetic EOFs (EOFs on  $\mathbf{S}_{multi}^{(col)}$ ). (Top) Factor maps for the first two dimensions. (Bottom) Loadings of each species for the first two dimensions. For each dimension, the loadings of the species represent when the several 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.

780 **S.1 Supplementary material**

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

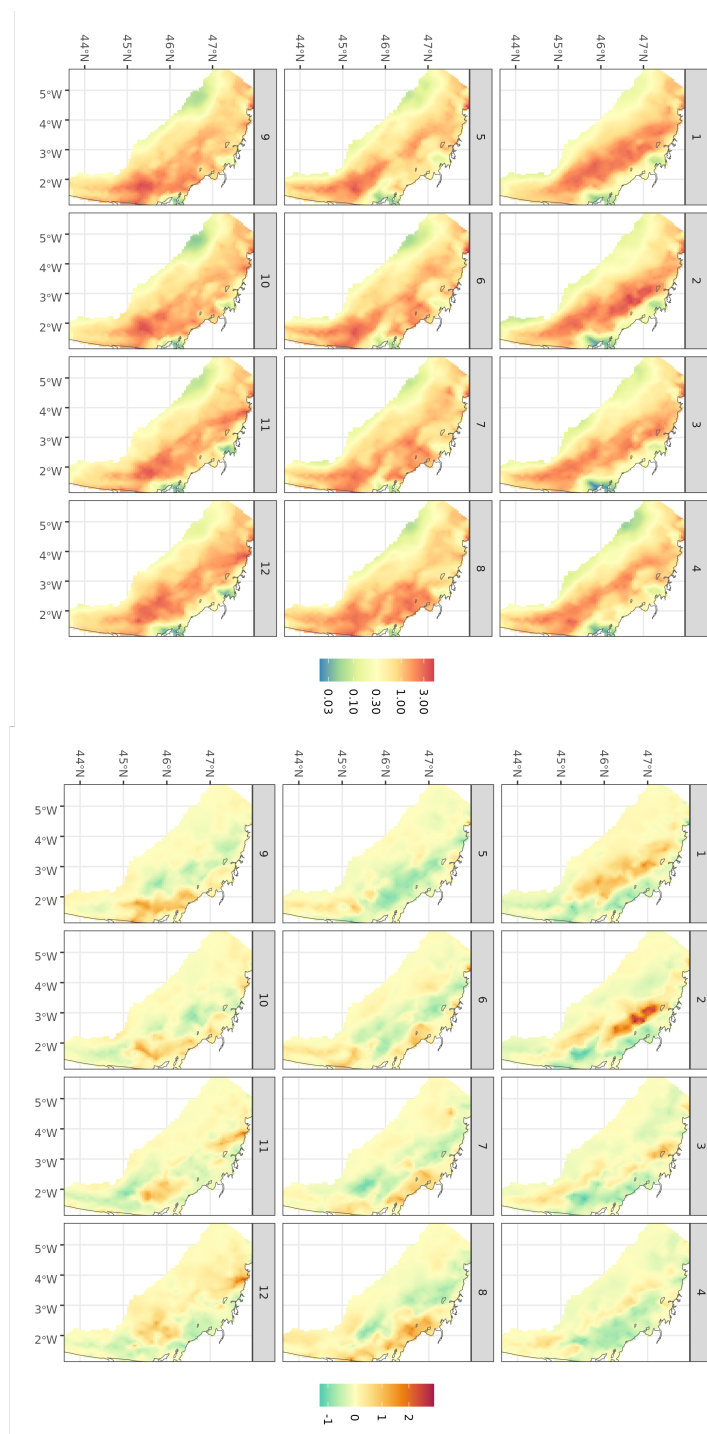


Fig. S1: IST-SDM predictions. (left) Monthly spatial predictions for sole in the log-scale  $\log S(x, t)$ . (right) Monthly anomalies of the spatial predictions  $S^*(x, t)$ . Each panel corresponds to the average anomalies for a month over the period 2008 - 2018.

782 **S.1.2 Additional dimensions for EOFs**

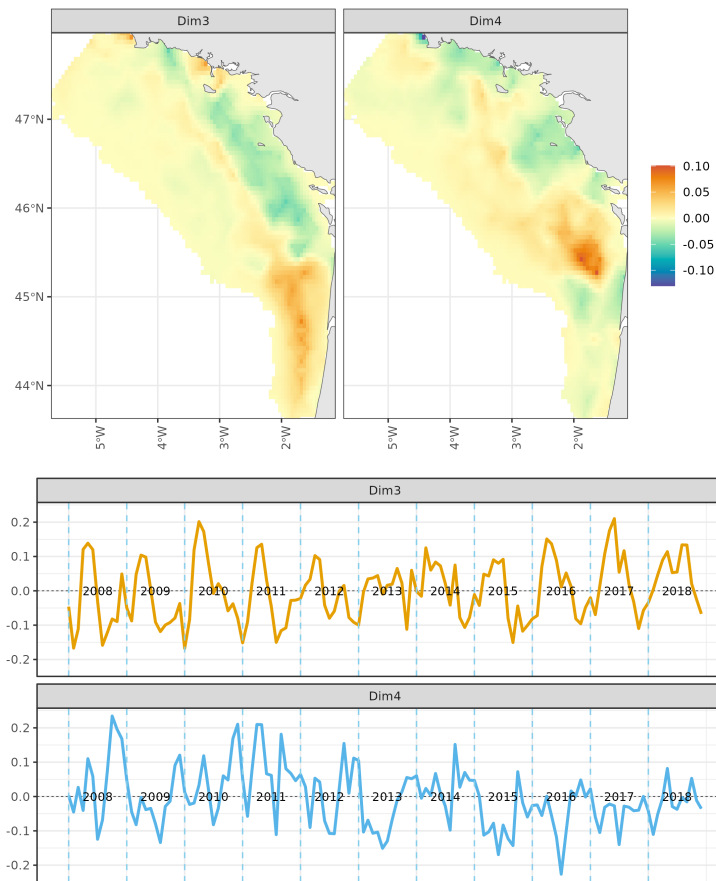


Fig. S2: IST-SDM predictions. (Top) Spatial factors for the dimensions three and four of the EOFs. (Bottom) Loadings for dimensions three and four of the EOFs. Blue dashed vertical lines correspond to the month of January for each year.

783 **S.1.3 CCA for the IST-SDM case study**

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



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

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

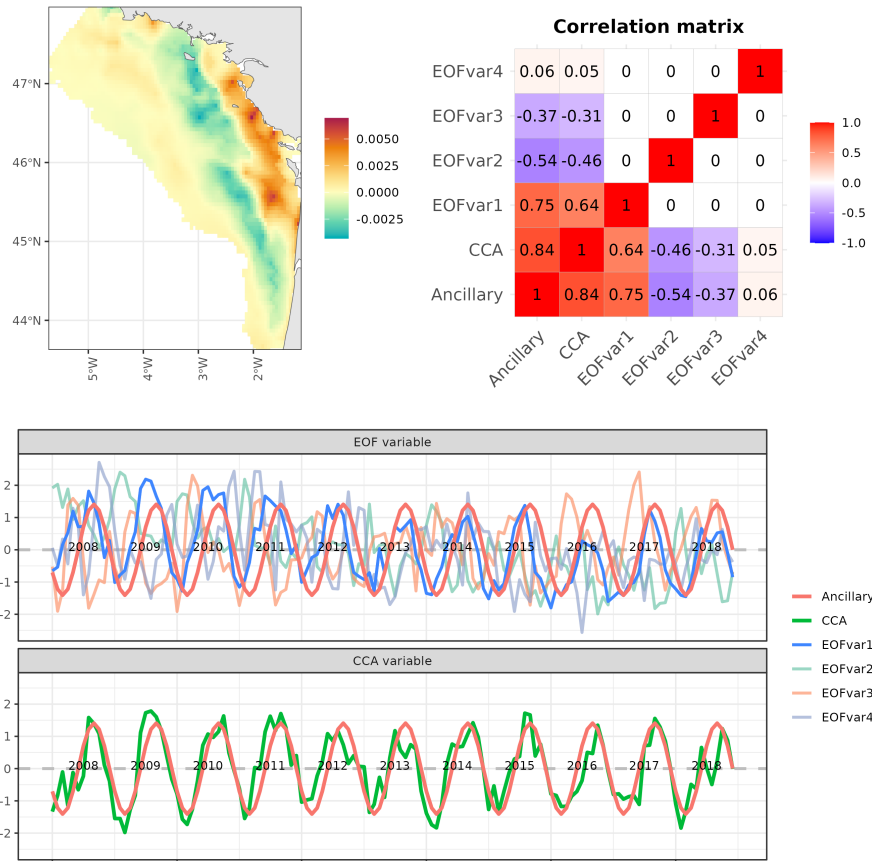


Fig. S3: IST-SDM predictions. Results of the canonical correlation analysis. (Top left) Spatial basis vectors  $\mathbf{w}_1$  that maximize the correlation between the temporal variables  $\mathbf{y}_1$  and  $\mathbf{y}_2$ . (Top right) Correlation matrix between the four first loading factors of the 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.

797 **S.1.4 Lag EOFs on satellite NDVI data**

798 **S.1.4.1 Data and matrix structure**

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

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

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

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

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

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

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

806 full year.  $\tau = \{0, 1, 2, \dots, 11\}$  would evidence sequence with monthly lags over a full year.

807 The matrix  $\mathbf{S}'_{lag}$  is then diagonalized like standard EOFs.

808 In this case, the spatial factors ( $\mathbf{u}_1$  for instance) are composed of  $\mathcal{T}$  maps and the temporal

809 loadings (*e.g.*  $\mathbf{v}_1$ ) are vectors of length  $p - \mathcal{T}$ .

### 810 **S.1.5 Illustration with NDVI data**

811 We take NDVI satellite data to illustrate lag EOFs. In the example, we consider quarterly

812 lags, so  $\tau = \{0, 3, 6, 9\}$ . Based on the scree plot (Figure 2, top right), we select the three

813 first dimensions that represents respectively 20%, 20% and 10% of the variance.

814 The first and second dimension have an annual period while the third one have a six-month

815 period. This last cycle was not evidenced by the standard EOF.

816 First and second dimension of the lag EOFs are very similar to the one of the standard EOFs

817 with a 3-months offset. They outline the annual variation of NDVI throughout the year with

818 high values of NDVI in summer specifically in mountains and low values during winter.

819 In the third dimension, loadings are high in February/March as well as August/September

820 and low in May/June as well as November/December (Figure S5). The areas that contribute

821 most to the sequence of maps are areas in the North of France (Figure S6). This dimension

822 highlights that there are peaks of vegetation in spring (April/May) and later in autumn

823 (September/October) that can be related to the higher growth of vegetation due most likely

824 to rain and moderate temperatures in these areas at this period of the year. By contrast,

825 NDVI is low in July and November which are either dry periods for summer or cold period

826 which tend to reduce the growth of vegetation.

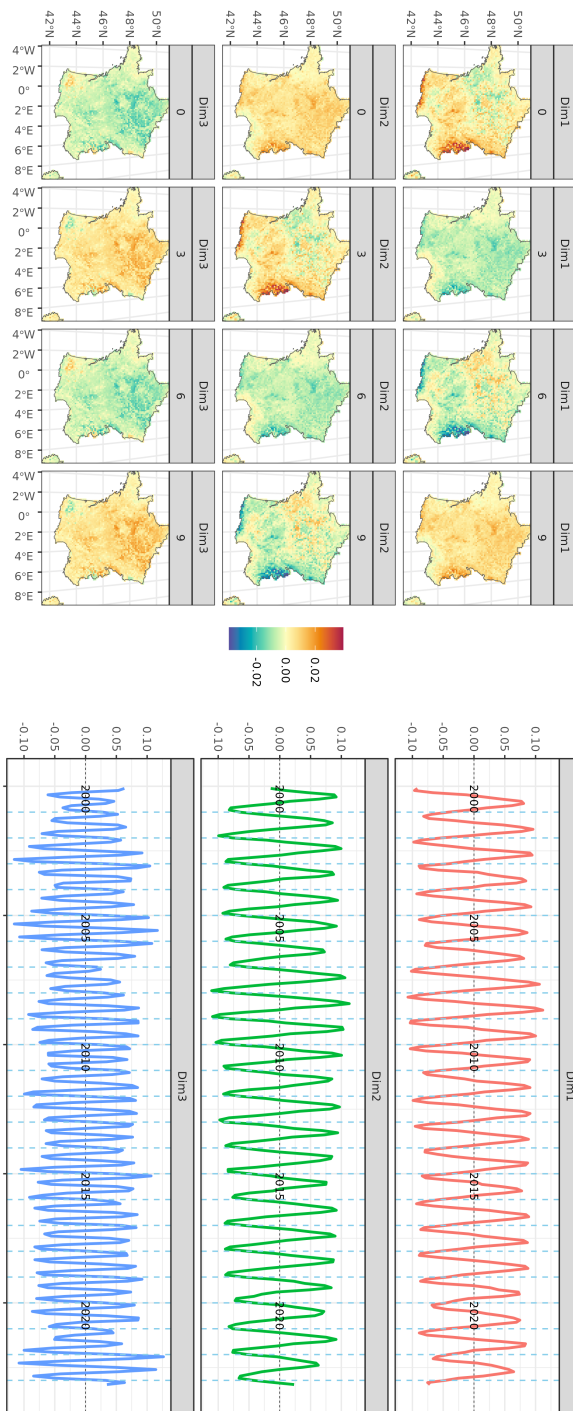


Fig. S4: Satellite NDVI data. (Top) Spatial factors for the three first dimensions of the lag EOFs. (Bottom) Loadings for the three first dimensions of the lag EOFs. Blue dashed vertical lines correspond to the month of January for each year.

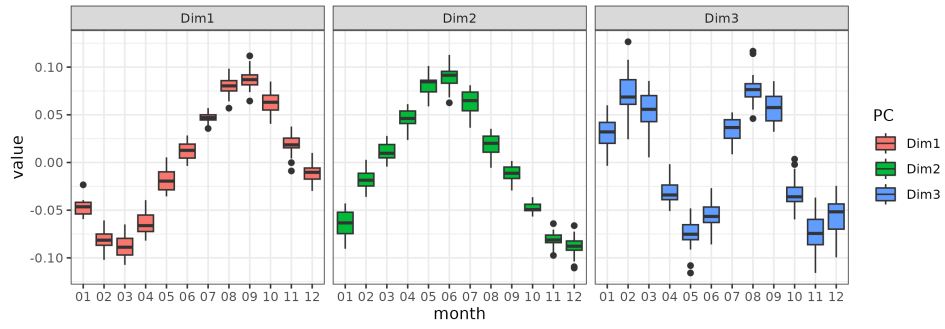


Fig. S5: Satellite NDVI data. Monthly boxplot of the temporal loadings.

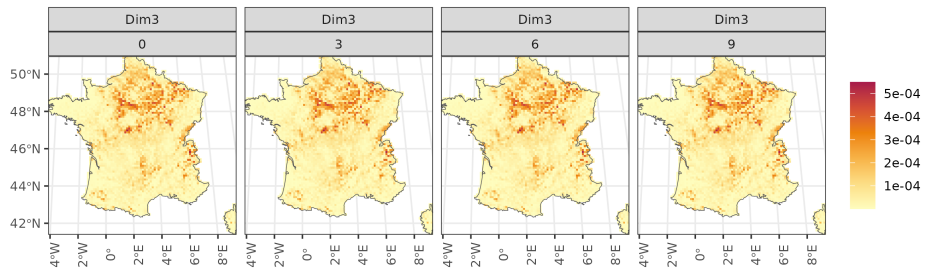


Fig. S6: Satellite NDVI data. Local variance explained by the lag EOFs for the third dimension.

827 **S.1.6 Multispecies analysis**

828 **S.1.6.1 EOFs raw results**

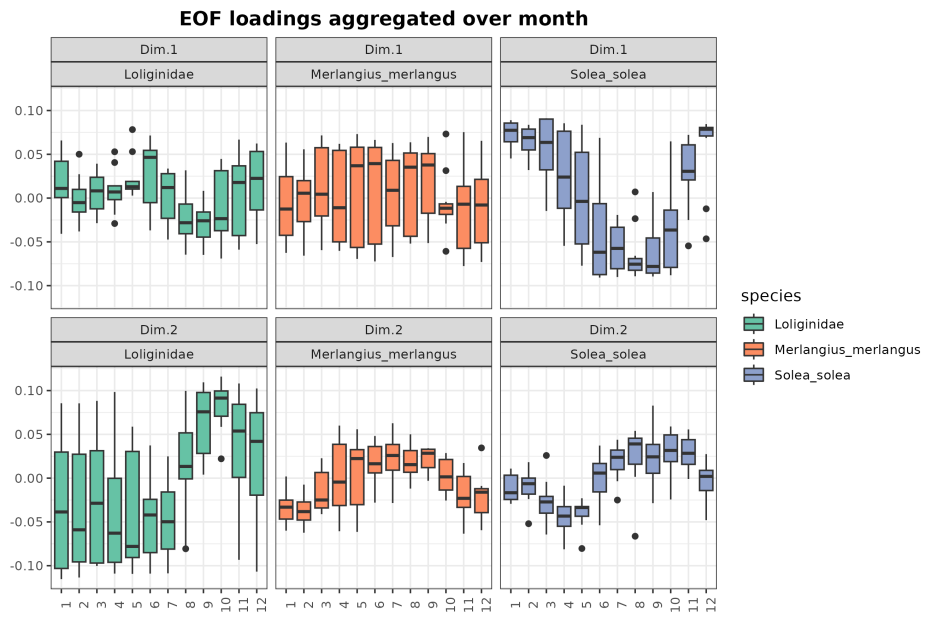


Fig. S7: IST-SDM predictions. Loadings of the multivariate analysis (Figure 6) aggregated over months.

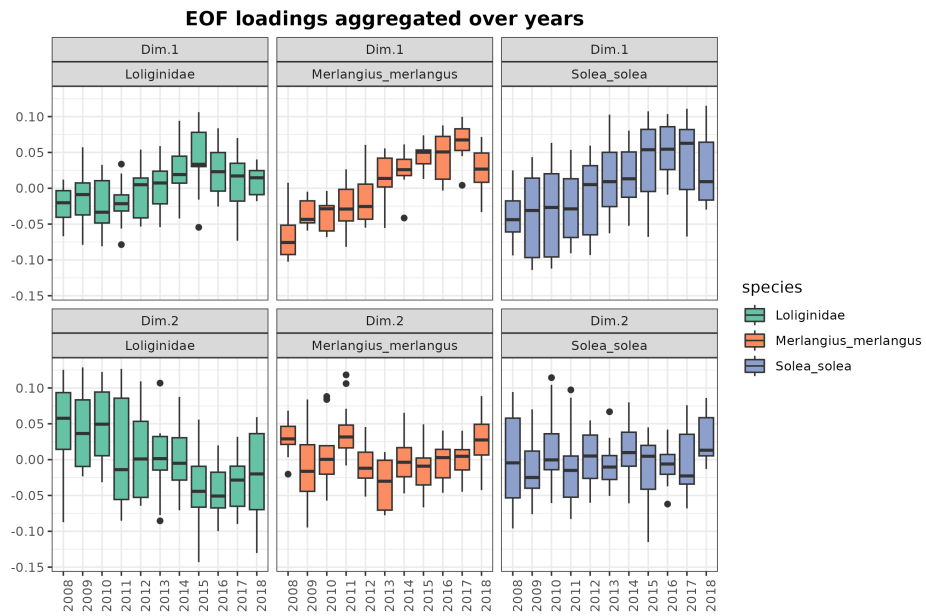


Fig. S8: IST-SDM predictions. Loadings of the multivariate analysis aggregated over years.



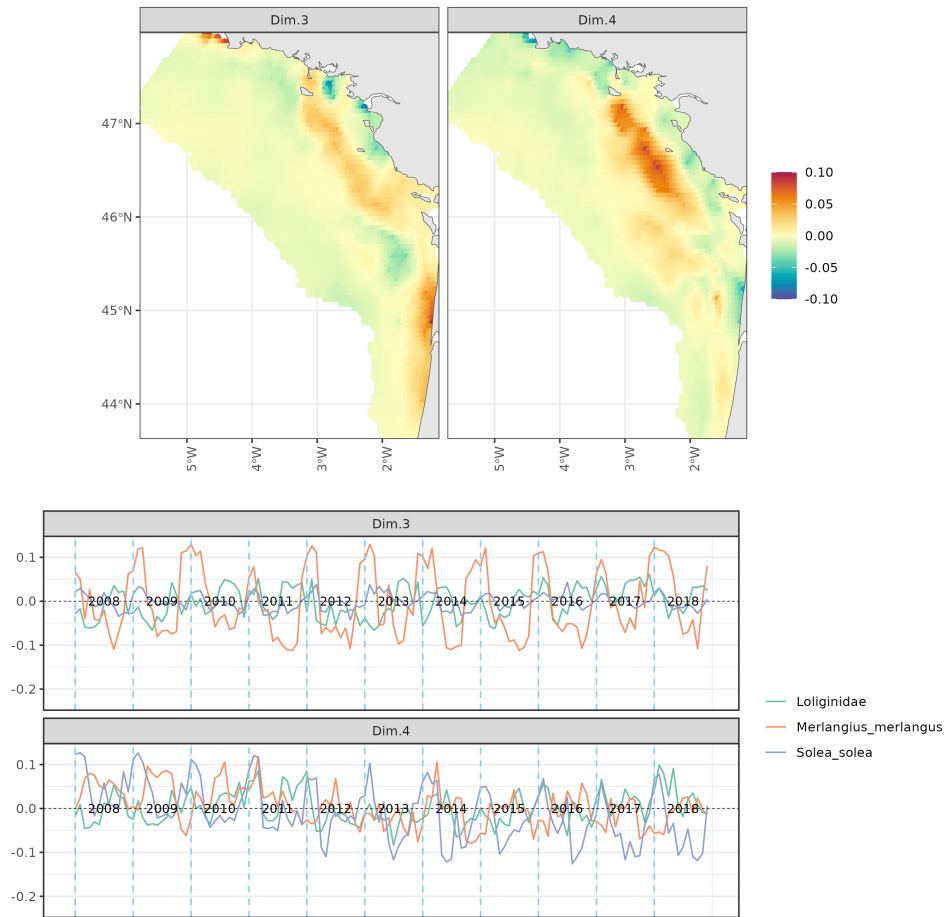


Fig. S9: IST-SDM predictions. Multivariate EOFs on the matrix  $\mathbf{S}_{multi}^{(col)}$ . (Top) Spatial factors for dimensions three and four. (Bottom) Loadings of each species for dimensions three and four. For each dimension, the loadings of the species represent when the several species of the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

### 829 **S.1.6.2 Clustering analysis results**

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

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

840 In a multivariate perspective, depending whether we choose the temporal synthetic repre-  
841 sentation or the spatially synthetic representation, the information that comes out can be  
842 different.

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

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

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

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

#### 855 **Temporally synthetic representation**

856 The spatial clusters 3 and 5 are related to seasons B and D which are mainly winter and  
857 spring months. These are reproduction period for the species under study (Arbault, Camus,  
858 and Bec, 1986; Houise, 1993; Moreno et al., 2002); the related spatial clusters can be  
859 interpreted as reproduction areas for these species. This is consistent with the previous  
860 studies investigating the reproduction areas of these species (Alglave, Vermard, et al.,  
861 2023). Interestingly, the spatial cluster 5 is only found for whiting and highlights a specific  
862 reproduction dynamics in the North of the Bay of Biscay during winter ( $4.5^{\circ}\text{W} - 4^{\circ}\text{N}$ ).

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

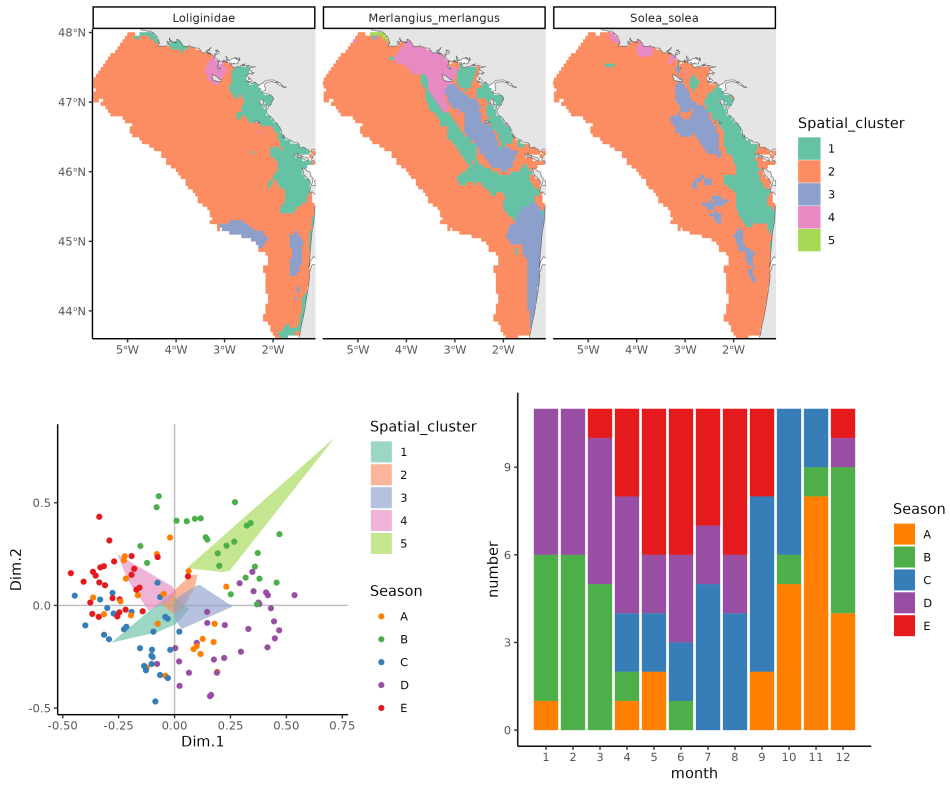


Fig. S10: Clustering analysis for the EOFs outputs of the temporally synthetic representation (EOFs on  $S^{(row)}_{multi}$ ). (Top row) Spatial clusters representing the functional zones in space for each three species. (Bottom left) Temporal loadings and spatial clusters projected in the two first 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).

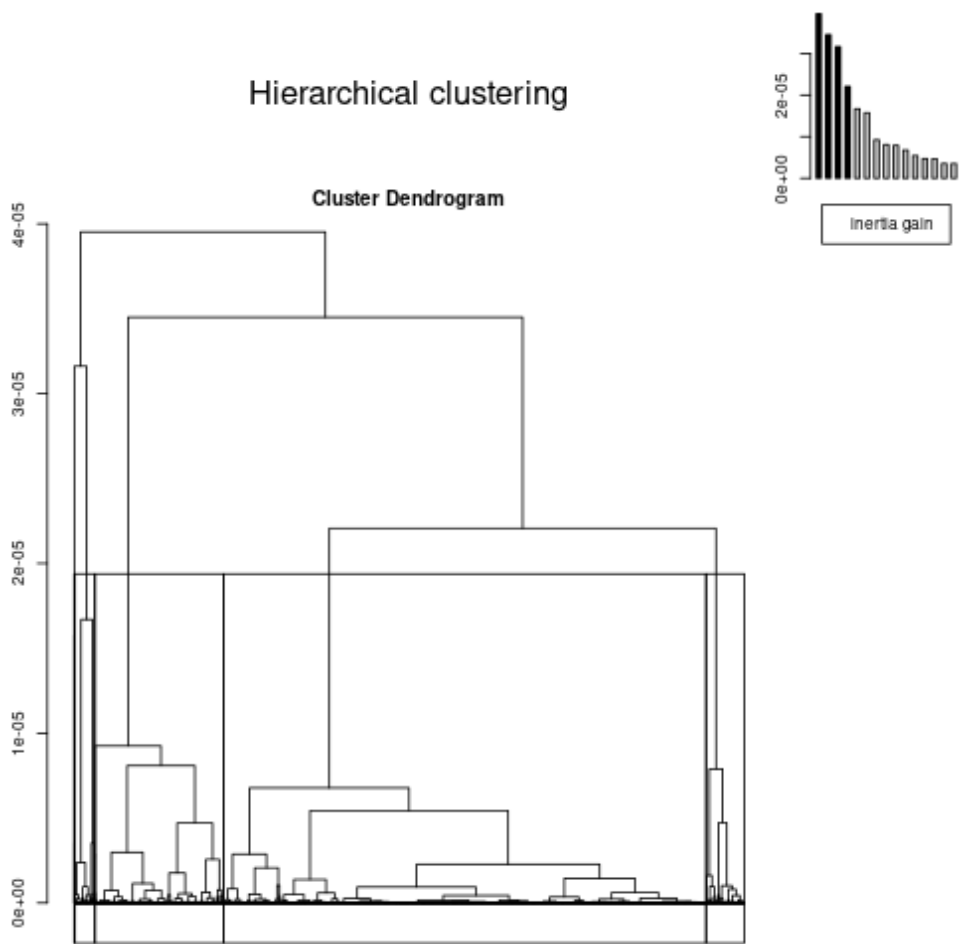


Fig. S11: Clustering tree of the spatial locations for the EOFs outputs of the temporally synthetic representation (EOFs on  $S_{multi}^{(row)}$ ).

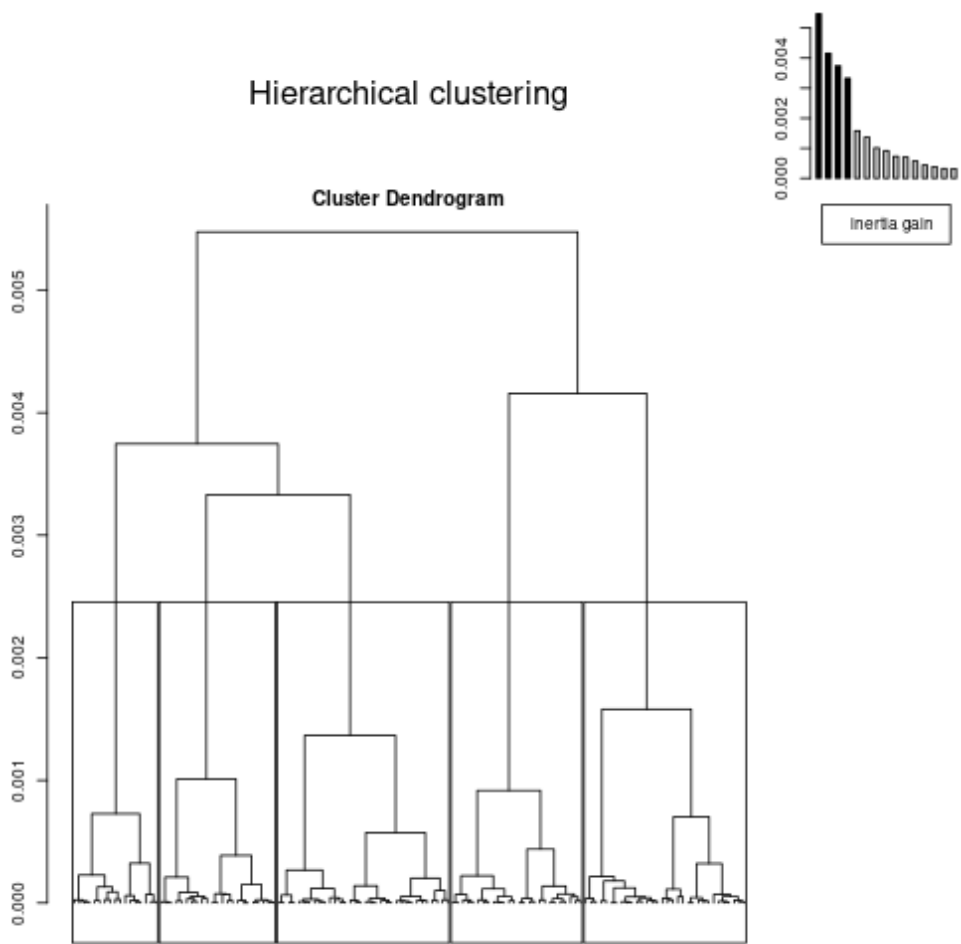


Fig. S12: Clustering tree of the temporal loadings for the EOFs outputs of the temporally synthetic representation (EOFs on  $S_{multi}^{(row)}$ ).

867 **Spatially synthetic representation**

868 Five spatial clusters are identified through the spatially synthetic representation of the EOFs.

869 The spatial cluster 1 is related to season B that fall mainly in summer and autumn for each  
870 three species. The related functional area (spatial cluster 1) can be considered as a summer  
871 grounds for each three species.

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

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

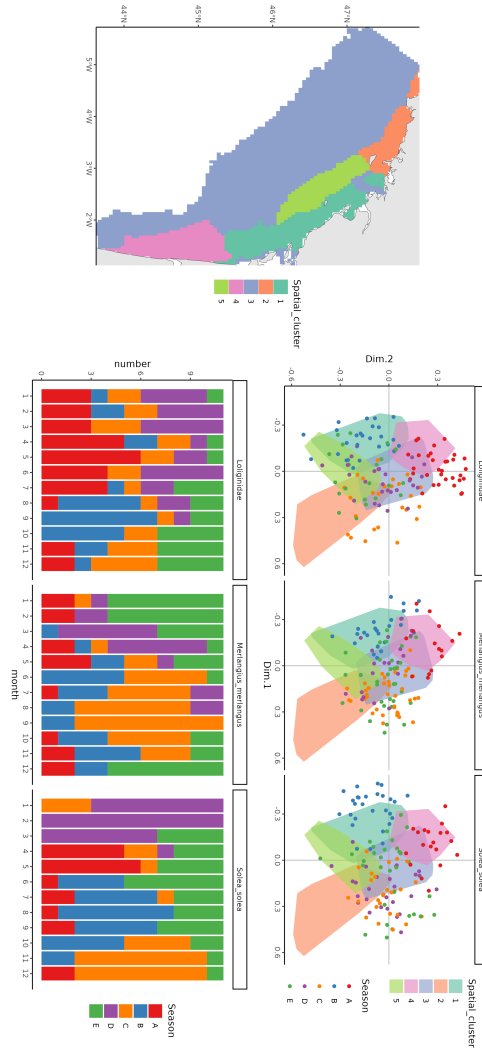


Fig. S13: Clustering analysis for the EOFs outputs of the spatially synthetic representation (EOFs on  $\mathbf{S}^{(col)}$ ). (Left) Spatial clusters representing the functional zones in space. (Top row) Projection of the temporal loadings and spatial locations in the two first dimensions of the PCA. 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).



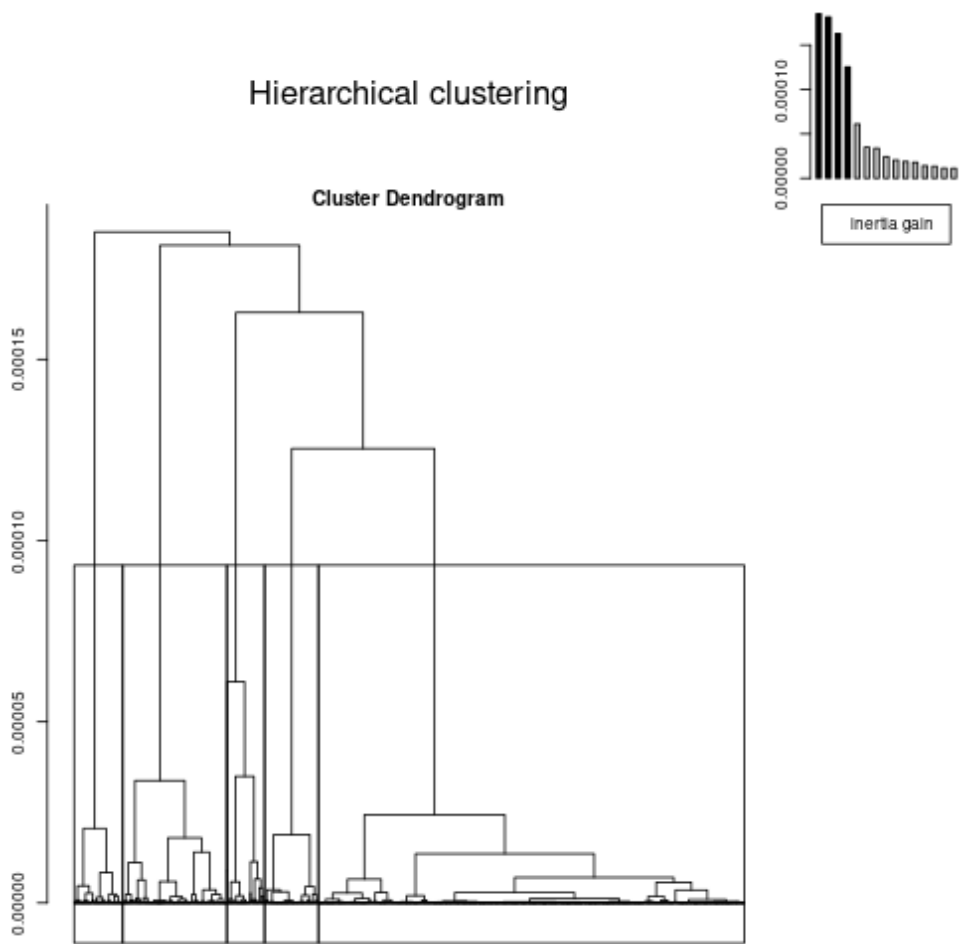


Fig. S14: Clustering tree of the spatial locations for the EOFs outputs of the spatially synthetic representation (EOFs on  $S_{multi}^{(col)}$ ).

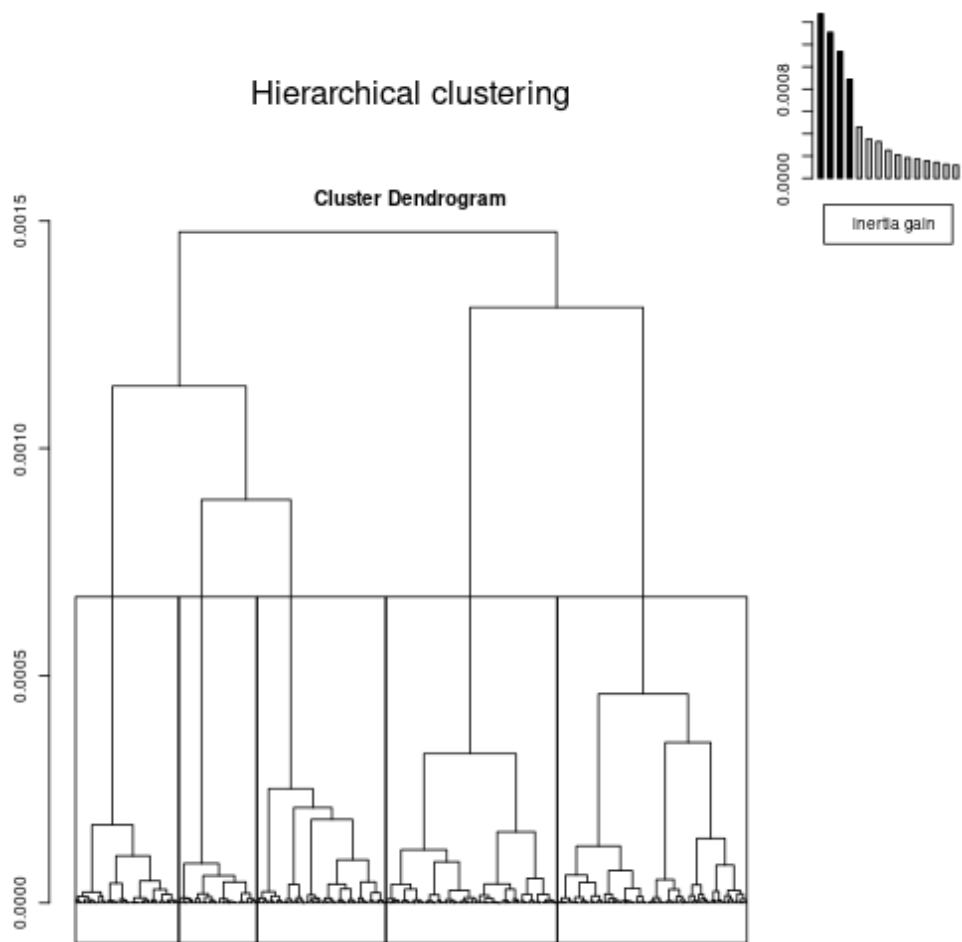


Fig. S15: Clustering tree of the temporal loadings for the EOFs outputs of the spatially synthetic representation (EOFs on  $S_{multi}^{(col)}$ ).