Supplementary Material 4

Additional results

S4.1 Percentage of variance explained

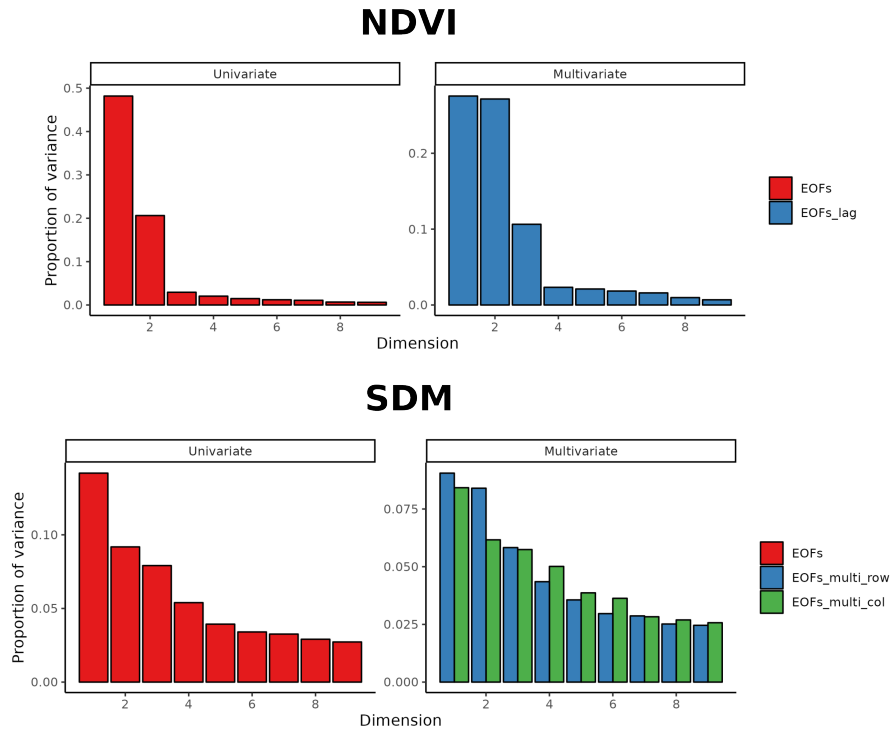

Fig. S4.1: Proportion of variance explained by each method for each case study. Top: NDVI satellite data, bottom: IST-SDM predictions. (Left) Univariate analysis. (Right) Multivariate analysis. EOFs_lag is the lagged version of EOFs. EOFs_multi_row is the multivariate EOFs conducted on $\mathbf{S}_{multi}^{(row)}$ (temporally synthetic representation). EOFs_multi_col is the multivariate EOFs conducted on $\mathbf{S}_{multi}^{(col)}$ (spatially synthetic representation).

S4.2 Analysis of the EOFs results

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

By contrast, the second dimension emphasizes a slightly different phase of seasonality. Loadings are negative from March to June (spring and early summer), which corresponds to high NDVI in the western and northern part of France. This can be related to both rain and higher temperature conducting to high primary production in this area. From July to November, loadings are positive. NDVI is high in mountains (similar to the first dimension) and low in the western and northern part of France. Note that nothing guarantees these two phases are not a signal arising from the same ecological process; one could possibly constrain these dimensions to match some temporal ancillary variable representing the common underlying process.

For the IST-SDM case study, the first dimension highlights strong seasonality. The loadings exhibit highly positive values in winter (January to February) and strongly negative values during summer (April to August), indicating that soles are concentrated in red/orange areas during winter and in blue areas during summer. The second dimension also emphasizes the same seasonality but primarily depicts a long-term decreasing trend. It suggests that fish progressively migrate northward over the period and become more concentrated in Northern red areas. Note that dimension 3 also emphasizes seasonal patterns (Figure S4.2) related to reproduction and dimension 4 highlights an important hotspot of sole distribution in the south of the Bay of Biscay (2°W - 45.5°N).

S4.3 Additionnal dimensions for EOFs

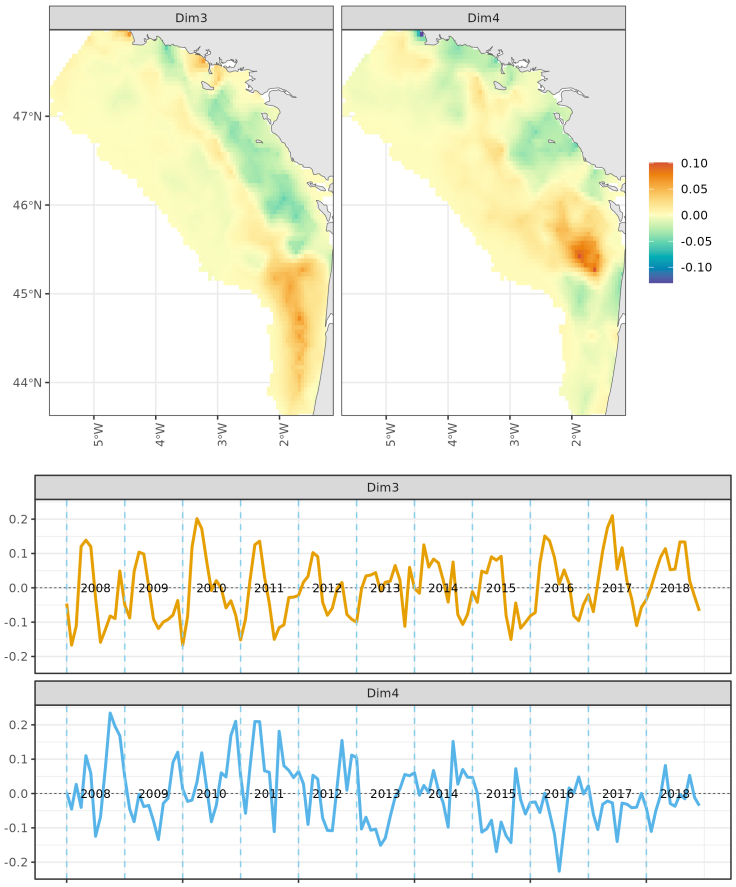

Fig. S4.2: IST-SDM predictions. (Top) Factors for the dimensions three and four of the EOFs. (Bottom) Loadings for dimensions three and four of the EOFs. Blue dashed vertical lines correspond to the month of January for each year.

S4.4 CCA for the IST-SDM case study

For IST-SDM predictions, we use the same parameterisation for the ancillary variable thought the delay D is fixed to $D = 0$ so that the peak of the

seasonal variable falls in March (the peak of the reproduction period). This way we aim at disentangling the seasonnal signal related to the alternance of reproduction/feeding that can be found in the four first dimensions.

The top-left panel of Figure S4.3 shows the resulting spatial basis vector, which reinforces the offshore-onshore gradient observed in the EOFs results. The inclusion of the ancillary variable enhances the clarity of these spatially distinct positive and negative regions within the canonical vector. The first three EOFs show a positive correlation with the ancillary variable, with the highest correlation coefficient being observed between CCA and the ancillary variable. This indicates that the CCA effectively extracts the temporal variation from the first three dimensions, resulting in a new variable that best aligns with the ancillary variable.

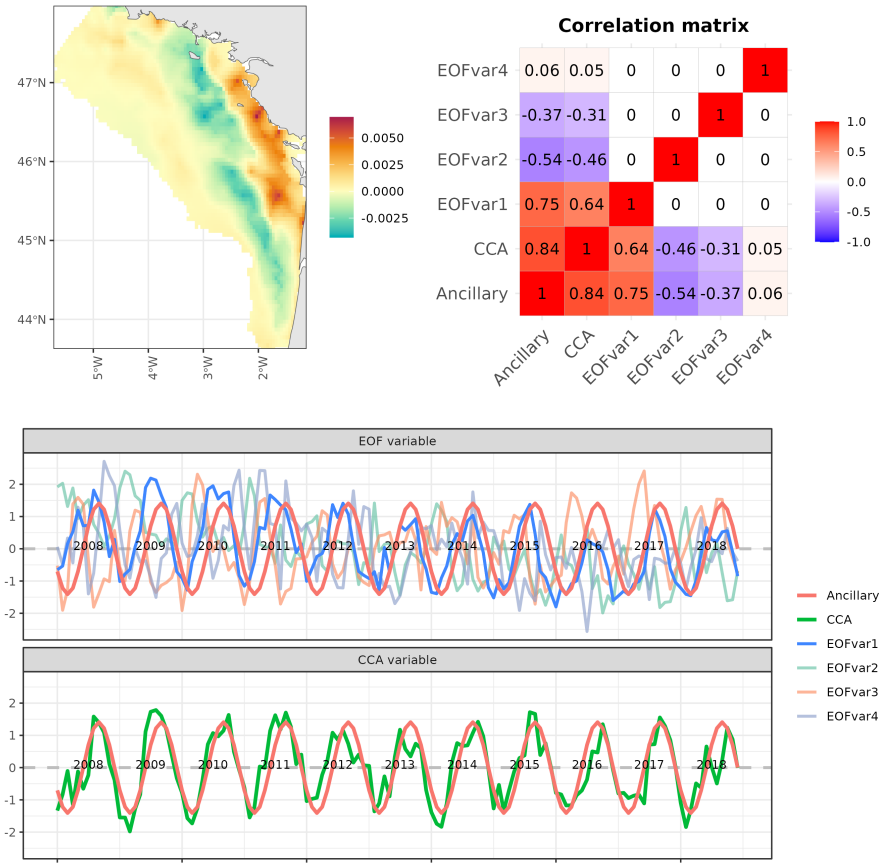

Fig. S4.3: IST-SDM predictions. Results of the canonical correlation analysis. (Top left) Spatial basis vectors $\mathbf{w}_1$ that maximize the correlation between the temporal variables $\mathbf{y}_1$ and $\mathbf{y}_2$. (Top right) Correlation matrix between the four first loading factors of the EOFs, 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.

S4.5 Spatially synthetic multivariate EOFs

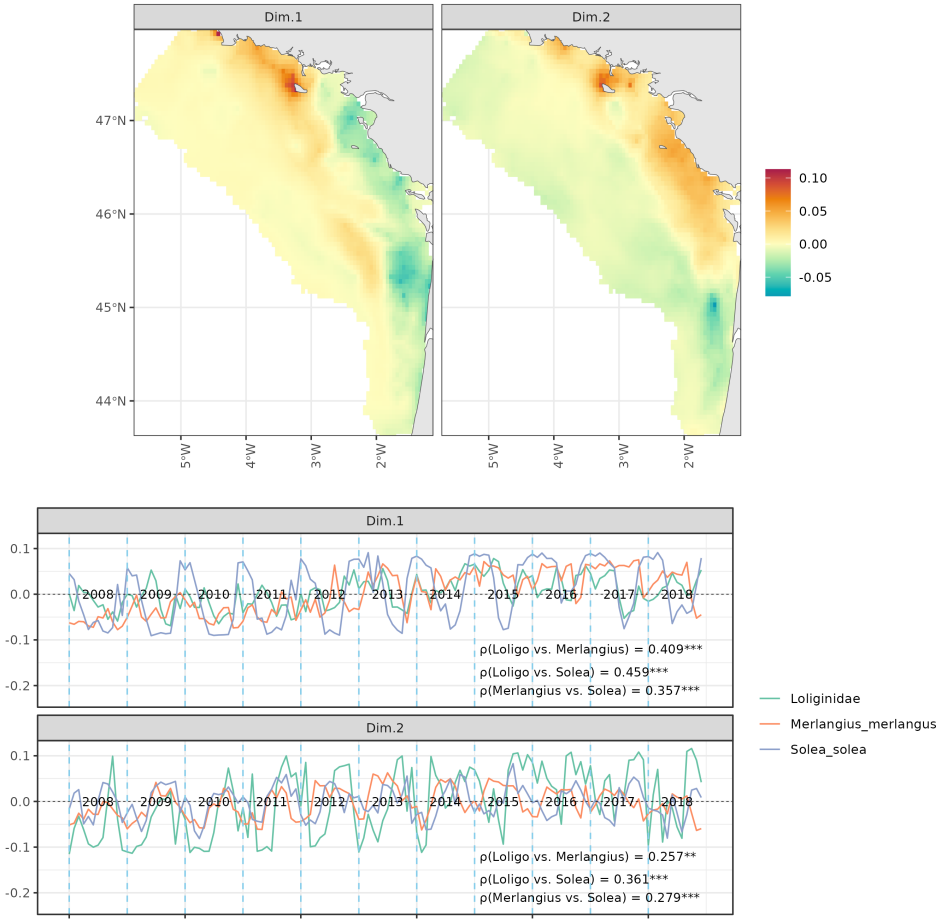

Fig. S4.4: IST-SDM predictions. Results for multivariate spatially synthetic EOFs (EOFs on $\mathbf{S}'^{(col)}_{multi}$). (Top) Factor maps for the first two dimensions. (Bottom) Loadings of each species for the first two dimensions. Blue dashed vertical lines correspond to the month of January for 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 dimensions of the EOFs.

The spatially synthetic multivariate EOFs (EOFs on $\mathbf{S}_{multi}^{\prime(col)}$) provides one single spatial pattern per dimension and as many loadings as variables. An angle appears after the fourth dimension of the scree plot (Figure S4.1). Interestingly, the first factor gathers the main reproduction areas for sole and squids that were previously identified (Figure S4.4 and S4.8). The second dimension emphasizes an overall increase in the loadings, illustrating a general drift towards the North of the Bay of Biscay and in the offshore areas (Figure S4.9). The loadings are correlated among species, highlighting that they share common temporal dynamics relatively to the factors of these dimensions (Figure S4.4).

S4.6 Lag EOFs on satellite NDVI data

S4.6.1 Data and matrix structure

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

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

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

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

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

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

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

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

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

S4.6.2 Illustration with NDVI data

We take NDVI satellite data to illustrate lag EOFs. In the example, we consider quarterly lags, so $\boldsymbol{\tau} = \{0, 3, 6, 9\}$. Based on the scree plot (Figure S4.1, top right), we select the three first dimensions that represents respectively 20%, 20% and 10% of the variance.

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

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

In the third dimension, loadings are high in February/March as well as August/September and low in May/June as well as November/December (Figure S4.6). The areas that contribute most to the sequence of maps are

areas in the North of France (Figure S4.7). This dimension highlights that there are peaks of vegetation in spring (April/May) and later in autumn (September/October) that can be related to the higher growth of vegetation due most likely to rain and moderate temperatures in these areas at this period of the year. By contrast, NDVI is low in July and November which are either dry periods for summer or cold period which tend to reduce the growth of vegetation.

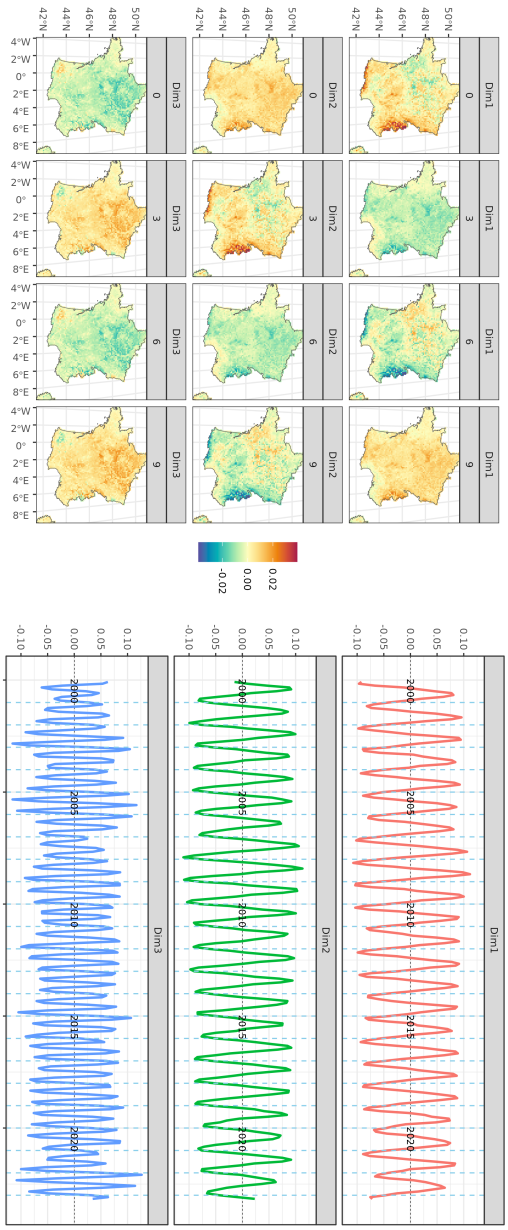

Fig. S4.5: Satellite NDVI data. (Top) Factors for the three first dimensions of the lag EOFs. (Bottom) Loadings for the three first dimensions of the lag EOFs. Blue dashed vertical lines correspond to the month of January for each year.

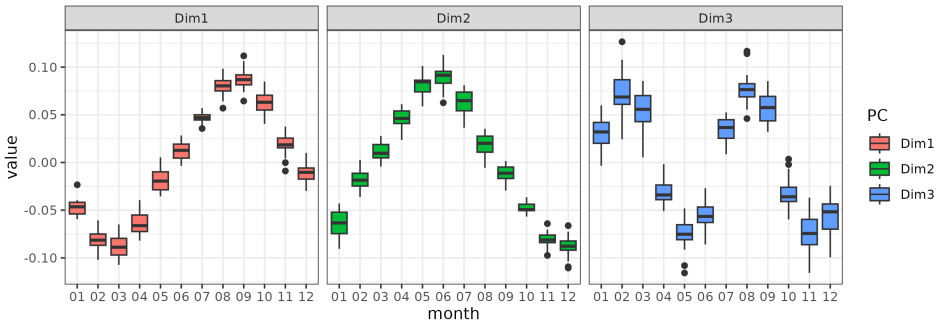

Fig. S4.6: Satellite NDVI data. Monthly boxplot of the loadings.

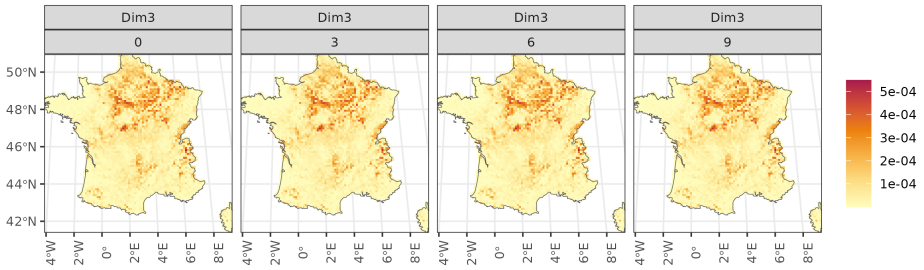

Fig. S4.7: Satellite NDVI data. Local variance explained by the lag EOFs for the third dimension.

S4.7 Multispecies analysis

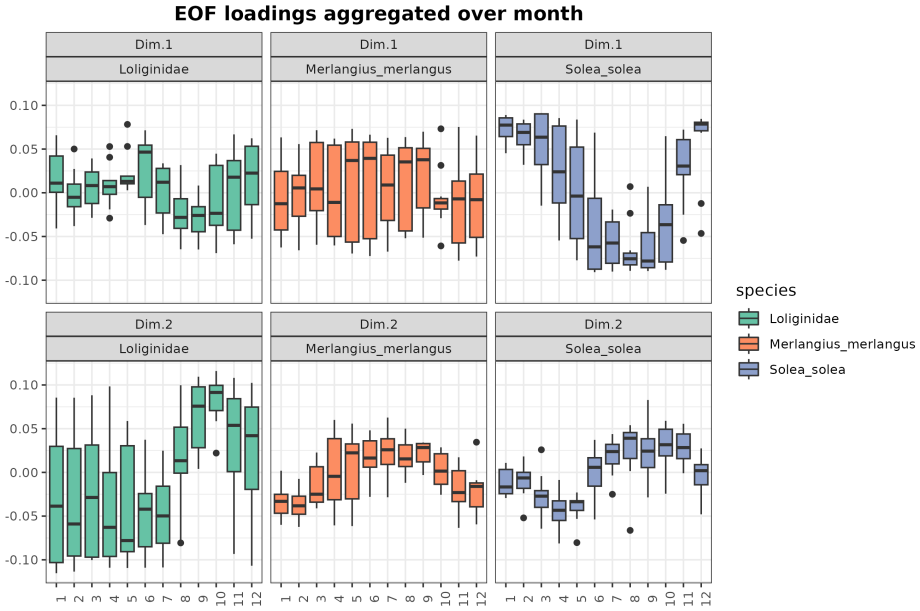

Fig. S4.8: IST-SDM predictions. Loadings of the multivariate analysis (Figure S4.4) aggregated over months.

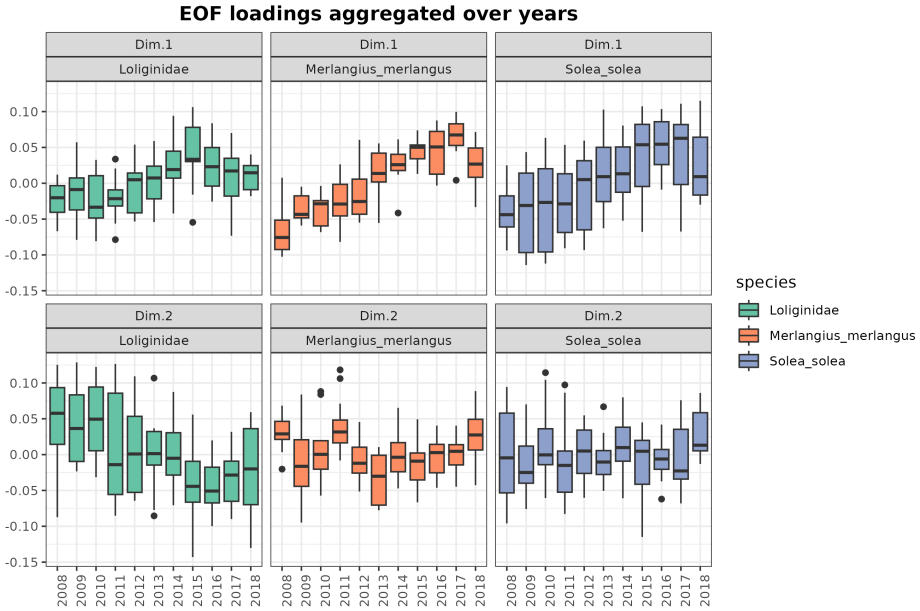

Fig. S4.9: IST-SDM predictions. Loadings of the multivariate analysis aggregated over years.

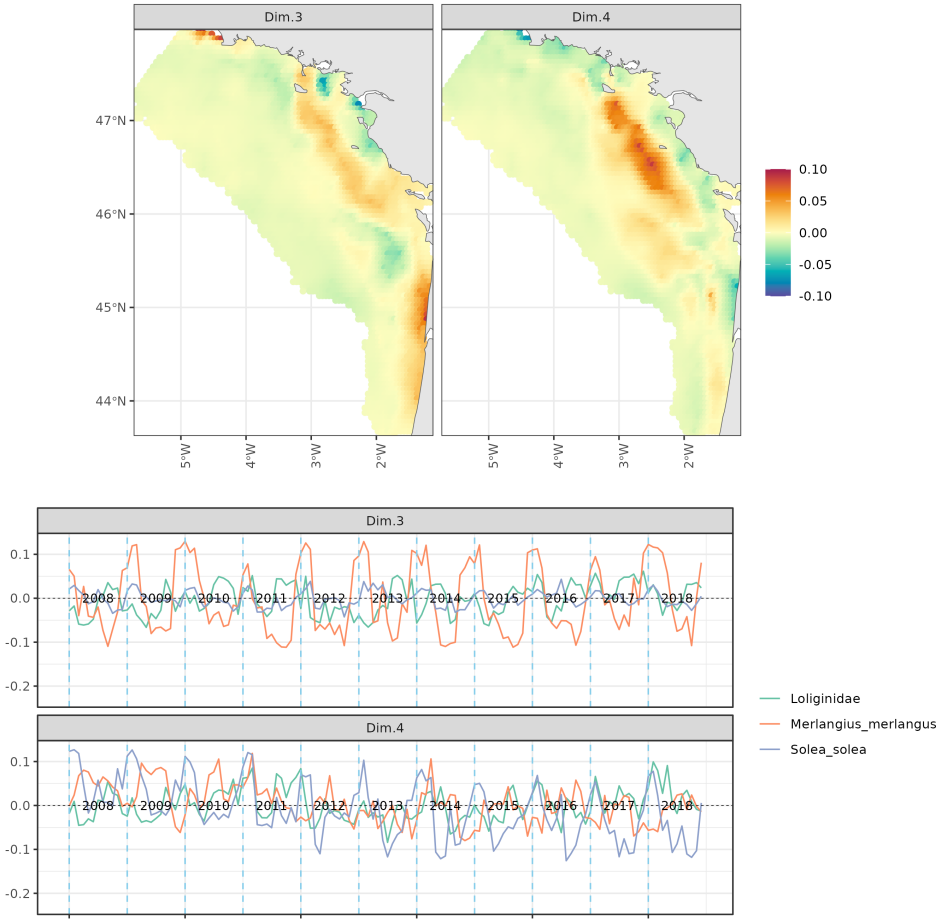

Fig. S4.10: IST-SDM predictions. Multivariate EOFs on the matrix $\mathbf{S}'^{(col)}_{multi}$. (Top) Factors for dimensions three and four. (Bottom) Loadings of each species for dimensions three and four. For each dimension, the loadings of the species represent when the several species of the analysis follow the related factor. Blue dashed vertical lines correspond to the month of January for each year.

References

Weare, Bryan C and John S Nasstrom (1982). "Examples of extended empirical orthogonal function analyses". In: Monthly Weather Review 110.6, pp. 481–485.