Decomposing fish species distribution to identify characteristic spatial patterns

First Author[1] & Second author[1,2]

[1] Institution 1

[2] Institution 2

Author Note

Add authornote

Correspondence concerning this article should be addressed to First Author, Postal address. E-mail: my@email.com

Abstract

Here goes the abstract

Keywords: keywords

Word count: X

Decomposing fish species distribution to identify characteristic spatial patterns

1 Methods

1.1 Species, data and model outputs

1.1.1 Case studies.

1.1.2 Data.

1.1.3 Model and model outputs format.

1.2 Decomposing model outputs

1.2.1 EOF. Empirical Orthogonal Functions is a method that has been developped by Lorenz (1956) for weather forecasting applications. The original aim of the technic was to deduce from a set of spatio-temporal maps a smaller set of maps that best describe and summarise the spatio-temporal process of interest. The main idea was to define the spatio-temporal process $S(x,t)$ as a linear combination of spatial patterns p_m (named EOF) each related to temporal indices $\alpha_m(t)$ (Equation 1 - $x \in [\![1,N]\!], t \in [\![1,T]\!]$). These temporal indices indicate when the spatio-temporal process is distributed following their related spatial pattern.

$$S(x,t) = \sum_{m=1}^{M} \alpha_m(t) \cdot p_m(x) + r^M(x,t) \quad (1)$$

where $r^M(x,t)$ is the residual variation not captured by the M modes of variation $\alpha_m(t) \cdot p(m,x)$.

To estimate the temporal components $\alpha_m(t)$ and the spatial patterns $p(m,x)$, some criteria need to be set. A natural choice is to minimize the residual variation to best capture the variability $S(x,t)$ (i.e. $R^M = \sum_{m=1}^{M}(r_m^M)^2$ is minimized) and set orthogonal constraints between the modes of variability so that each mode is the "best" representation of variability from a statistical point of view (Equations 2, 3). This last criteria is further discussed in the discussion.

$$\sum_{x=1}^{N} p_m(x) \cdot p_j(x) = \delta_{mj} \equiv \begin{cases} 1 & \text{if} \quad m = j \\ 0 & \text{if} \quad m \neq j \end{cases} \tag{2}$$

$$T \cdot \overline{p_m^* p_j^*} = a_m \delta_{mj} \tag{3}$$

with $a_m \geq a_{m+1} \geq 0$, $\overline{(\ )}$ denoting the time average and $(\ )^*$ a departure from the time average.

Let's write these equations in matrix terms by introducing the $T \times N$ matrix $\mathbf{S}$, $\mathbf{S}^*$, $\mathbf{Q}$, $\mathbf{Q}^*$. They respectively refer to $S(x,t)$, $S^*(x,t)$, $\alpha_m(t)$ and $\alpha_m^*(t)$. We denote $\mathbf{Y}$ is a square matrix of order N with elements $p_m(x)$ (NEED TO CHECK). Then, the problem can be reformulated as:

$$\mathbf{S} = \mathbf{Q}\mathbf{Y} \tag{4}$$

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{I} \tag{5}$$

$$\mathbf{Q}^{*T}\mathbf{Q} = \mathbf{D} \tag{6}$$

with $\overline{(\ )}$ the transpose, $\mathbf{I}$ the identity, $\mathbf{D}$ a diagonal matrix with diagonal being equal to a_m/T.

Finally, by introducing $A \equiv Q^{*T}Q^*$ (whose elements are proportionnal to the covariance of $\alpha_m(t)$), we can rewrite these equations as:

$$\mathbf{Y}\mathbf{A}\mathbf{Y}^T = \mathbf{D} \tag{7}$$

This is a standard “eigenvalue-eigenvector” problem. As a consequence, a link can be done with standard multivariate analysis such as PCA. Thus, in addition to the spatial patterns (and the related temporal components) that appear in the EOF formulations (Equation 1) and that can be obtained by diagonalysing the problem in Equation 7, additionnal analysis can be performed to obtain similar visualisation as in PCA (e.g. plot of individuals and variables) possibly coupled with standard clustering analysis.

1.2.2 Link with PCA. In this representation (Equation 7), $\mathbf{Y}^T = X$ is the matrix of $p_m(x)$ which are the eigen-vectors in PCA analysis, A is the covariance matrix of the centered predictions $\mathbf{S}^*$ (TO CHECK) and $\mathbf{D}$ is the diagonal matrix of the eigen-values.

Often, this equation cannot be resolved as the A matrix is not diagonalisable (when there are more spatial locations than time-steps). Then, it is required to pass through singular value decomposition to get the eigen-values and eigen-vectors of A.

Singluar value decomposition (svd) express a matrix $\mathbf{X}$ as a decomposition of 3 other matrices:

$\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}'$ with $\mathbf{U}\mathbf{U}' = \mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}$.

From these expressions, we can demonstrate that $\mathbf{X}\mathbf{X}' = \mathbf{V}'\boldsymbol{\Lambda}^2\mathbf{V}$ with $\boldsymbol{\Lambda}^2$ the eigen-values and V the eigen-vectors of the covariance matrix $\mathbf{X}\mathbf{X}'$. The eigen-values and eigen-vectors of the covariance matrix A where obtained from svd and allowed to performed the same analysis as in standard PCA (more precisely a PCA is a svd on centered or centered-reduced data if the PCA is normalized).

Then,

1.2.3 Representing variables and individuals. Spectral analysis: - https://bookdown.org/rdpeng/timeseriesbook/spectral-analysis.html - https://math.mcmaster.ca/~bolker/eeid/2010/Ecology/Spectral.pdf

2 Results

2.1 Percentage of variance explained by the EOF

–> 2 first capture around 20% each

2.2 Structuring spatial patterns and their temporal variation

–> Northward move for Hake and Sole

–> seasonnal PC, coincide with reproduction ecology –> result from migration patterns ?

2.3 Identying ecological seasons and related spatial patterns

–> result of clustering on temporal indices and related groups

–> related spatial patterns

3 Discussion

- Northward move –> climate change ?
- EOM ? Constraints to build the maps

–> "best" representation of variability from a statistical point of view = "second basis function should be spatially uncorrelated with the first, and this is equivalent to the requirement of spatial orthogonality"

(Lorenz, 1956; Randall, 2003)

(Monahan, Fyfe, Ambaum, Stephenson, & North, 2009)

4 References

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

Monahan, A. H., Fyfe, J. C., Ambaum, M. H., Stephenson, D. B., & North, G. R. (2009). Empirical orthogonal functions: The medium is the message. *Journal of Climate*, *22*(24), 6501–6514.

Randall, D. (2003). Empirical orthogonal functions. *Dept. Atmosph. Sci., Colorado State Univ., CO, USA*.