¹ Decomposing fish species distribution to identify characteristic spatial patterns

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

³ [1] Institution 1

⁴ [2] Institution 2

⁵ Author Note

⁶ Add authornote

⁷ Correspondence concerning this article should be addressed to First Author, Postal

⁸ address. E-mail: my@email.com

Abstract

Here goes the abstract

Keywords: keywords

Word count: X

13   Decomposing fish species distribution to identify characteristic spatial patterns

14   Species phenology –> major uncertainty

15   While key in understanding species life cycle and has major ecological and
16   conservation implications.

<h1 style="text-align:center">Methods</h1>

17

Species, data and model outputs

18

Case studies.   We selected sole and sea bass in the Bay of Biscay as well as Hake in
19
both the Bay of Biscay and the Celtic Sea as case studies. These three case studies are
20
selected as they all benefit from scientific and expert knowledge regarding their phenology
21
in order to validate the results of their analysis.
22

Sole of the Bay of Biscay reproduce.
23

Sea bass in the Bay of Biscay is .
24

Hake in the Bay of Biscay and in the Celtic Sea is considered as one single stock.
25
Still, there is question about the
26

To map species distribution, we filtered the data from several trawler fleets from 2008
27
to 2018.
28

Data.
29

Model and model outputs format.
30

Decomposing model outputs

31

EOF.   Empirical Orthogonal Functions is a method that has been developed by
32
Lorenz (1956) for weather forecasting applications. The original aim of the technic was to
33
deduce from a set of spatio-temporal maps a smaller set of maps that best describe and
34

35 summarise the spatio-temporal process of interest. The main idea was to define the

36 spatio-temporal process $S(x,t)$ as a linear combination of spatial patterns p_m (named

37 EOF) each related to temporal indices $\alpha_m(t)$ (Equation 1 - $x \in [\![1, N]\!], t \in [\![1, T]\!]$). These

38 temporal indices indicate when the spatio-temporal process is distributed following their

39 related spatial pattern.

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

40   where $r^M(x,t)$ is the residual variation not captured by the M modes of variation

41 $\alpha_m(t) \cdot p(m,x)$.

42   To estimate the temporal components $\alpha_m(t)$ and the spatial patterns $p(m,x)$, some

43 criteria need to be set. A natural choice is to minimize the residual variation to best

44 capture the variability $S(x,t)$ (i.e. $R^M = \sum_{m=1}^{M}(r_m^M)^2$ is minimized) and set orthogonal

45 constraints between the modes of variability so that each mode is the 'best' representation

46 of variability from a statistical point of view (Equations 2, 3). This last criteria is further

47 discussed in the discussion.

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

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

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

49 time average.

50   Let's write these equations in matrix terms by introducing the $T \times N$ matrix $\mathbf{S}$, $\mathbf{S}^*$,

51 $\mathbf{Q}$, $\mathbf{Q}^*$. They respectively refer to $S(x,t)$, $S^*(x,t)$, $\alpha_m(t)$ and $\alpha_m^*(t)$. We denote $\mathbf{Y}$ is a

52 square matrix of order N with elements $p_m(x)$ (NEED TO CHECK). Then, the problem

53 can be reformulated as:

$$\mathbf{S} = \mathbf{QY} \tag{4}$$

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

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

54   with $\overline{(\ )}$ the transpose, $\mathbf{I}$ the identity, $\mathbf{D}$ a diagonal matrix with diagonal being equal

55 to a_m/T.

56   Finally, by introducing $A \equiv Q^{*T}Q^*$ (whose elements are proportionnal to the

57 covariance of $\alpha_m(t)$), we can rewrite these equations as:

$$\mathbf{YAY}^T = \mathbf{D} \tag{7}$$

58   This is a standard "eigenvalue-eigenvector" problem. As a consequence, a link can be

59 done with standard multivariate analysis such as PCA. Thus, in addition to the spatial

60 patterns (and the related temporal components) that appear in the EOF formulations

61 (Equation 1) and that can be obtained by diagonalysing the problem in Equation 7,

62 additionnal analysis can be performed to obtain similar visualisation as in PCA (e.g. plot

63 of individuals and variables) possibly coupled with standard clustering analysis.

64   **Link with PCA.**   In this representation (Equation 7), $\mathbf{Y}^T = X$ is the matrix of

65 $p_m(x)$ which are the eigen-vectors in PCA analysis, A is the covariance matrix of the

66 centered predictions $\mathbf{S}^*$ (TO CHECK) and $\mathbf{D}$ is the diagonal matrix of the eigen-values.

67   Often, this equation cannot be resolved as the A matrix is not diagonalisable (when

68 there are more spatial locations than time-steps). Then, it is required to pass through

69 singular value decomposition to get the eigen-values and eigen-vectors of A.

⁷⁰   Singluar value decomposition (svd) express a matrix $\mathbf{X}$ as a decomposition of 3 other

⁷¹   matrices:

⁷² $\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}'$ with $\mathbf{UU}' = \mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{VV}' = \mathbf{V}'\mathbf{V} = \mathbf{I}$.

⁷³   From these expressions, we can demonstrate that $\mathbf{XX}' = \mathbf{V}'\boldsymbol{\Lambda}^2\mathbf{V}$ with $\boldsymbol{\Lambda}^2$ the

⁷⁴   eigen-values and $\mathbf{V}$ the eigen-vectors of the covariance matrix $\mathbf{XX}'$. The eigen-values and

⁷⁵   eigen-vectors of the covariance matrix $\mathbf{A}$ where obtained from svd and allowed to

⁷⁶   performed the same analysis as in standard PCA (more precisely a PCA is a svd on

⁷⁷   centered or centered-reduced data if the PCA is normalized).

⁷⁸   **Representing variables and individuals.**   Then, we analyse the spatio-temporal

⁷⁹   model outputs by presenting the spatial patterns $p_m(x)$ (i.e. the eigen-vectors) which are in

⁸⁰   the matrix $\mathbf{X} = \mathbf{V}$. These can be seen as maps that structure the actual process $\mathbf{S}$ (Cf.

⁸¹   Figure. . .) or alternatively as standard variable plot as in classical PCA analysis (Cf.

⁸²   Figure. . .).

⁸³   The temporal index $\alpha_m(t)$ are the projection matrix $\mathbf{S}^*$ onto the EOF basis function

⁸⁴ $\mathbf{V}$ i.e. $A = \mathbf{S}^*\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}$. These can be either seen as time series (Cf. Figure. . .) or as a

⁸⁵   classical plot of individuals on the first components of the PCA (Cf. Figure. . .).

⁸⁶   When confronted together, the spatial representations and the related temporal

⁸⁷   indices inform how the spatial patterns vary in time.

⁸⁸   **FIGURE: LINK BETWEEN EOF AND PCA**

⁸⁹   Spectral analysis: -

⁹⁰   https://bookdown.org/rdpeng/timeseriesbook/spectral-analysis.html -

⁹¹   https://math.mcmaster.ca/~bolker/eeid/2010/Ecology/Spectral.pdf

92   **Results**

93 **Strong seasonnal patterns**

94   –> 2 first capture around 20% each and often very seasonnal

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

96 patterns ?

97   –> Northward move for Hake and Sole

98 **Identifying seasons and functional zones**

99   –> result of clustering on temporal indices and related groups

100   High inter-annual stability, no confusion between month/cluster

101   –> related spatial patterns

102   **Discussion**

103   • Northward move –> climate change ?

104   • EOM ? Constraints to build the maps

105   –> 'best' representation of variability from a statistical point of view = 'second basis

106 function should be spatially uncorrelated with the first, and this is equivalent to the

107 requirement of spatial orthogonality'

108   (Lorenz, 1956; Randall, 2003)

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

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.*