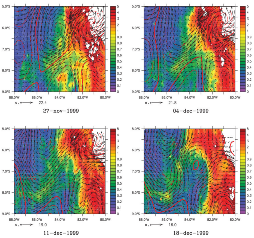
How best to ordinate spatio-temporal data?

Informing dimension-reduction with temporal ecological variables

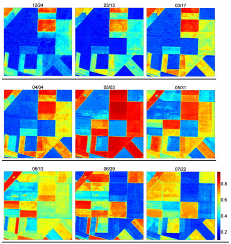
B. Alglave, S. Obakrim, B. Dufée and J. Thorson

July 2024

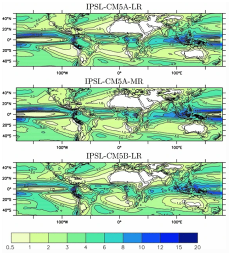
Oceanography



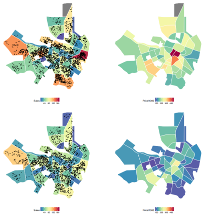
Agronomy



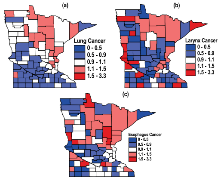
Climatology



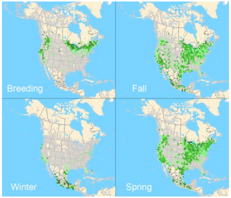
Economics



Epidemiology



Ecology



Empirical Orthogonal Functions

First introduced by Lorenz (1956)

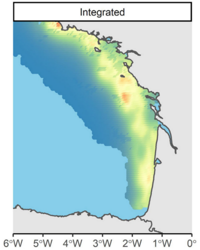
➠ What is the **best representation** for a spatio-temporal field ?

➠ How to perform **dimension-reduction** on a spatio-temporal field ?

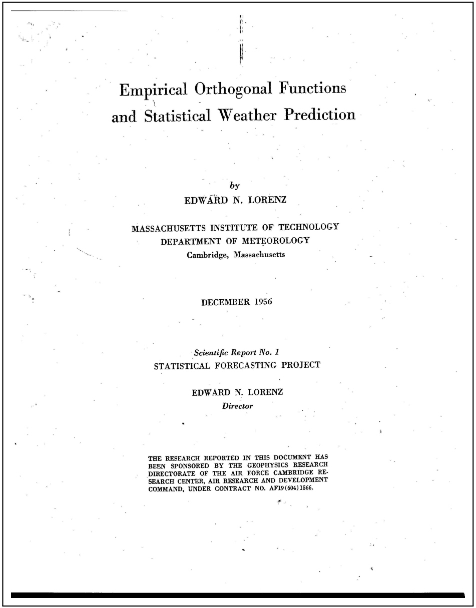
➠ How to make **projections**?

- -

Presentation based on a fishery case study.
Demersal fisheries of the Bay of Biscay.
Monthly maps from 2008 to 2018 (132 maps)



Solea solea



Empirical Orthogonal Functions
and Statistical Weather Prediction

by
EDWARD N. LORENZ

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF METEOROLOGY
Cambridge, Massachusetts

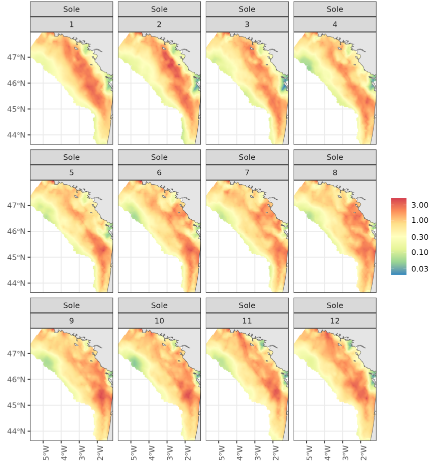
DECEMBER 1956

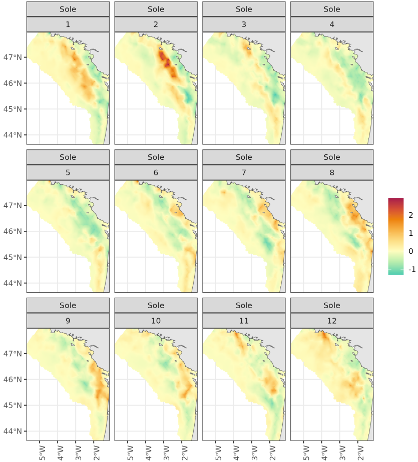
Scientific Report No. 1
STATISTICAL FORECASTING PROJECT

EDWARD N. LORENZ
Director

THE RESEARCH REPORTED IN THIS DOCUMENT HAS BEEN SPONSORED BY THE GEOPHYSICS RESEARCH DIRECTORATE OF THE AIR FORCE CAMBRIDGE RESEARCH CENTER, AIR RESEARCH AND DEVELOPMENT COMMAND, UNDER CONTRACT NO. AF19(604)1566.

Raw data





(left) Montly spatial log-predictions $\log S(x, t)$ of the hierarchical model. (right) Monthly anomalies of the spatial predictions $S^*(x, t)$. Each panel corresponds to the average distribution of prediction of anomalies for a month over the period 2008 - 2018.

Basics of EOF

The spatio-temporal field is decomposed so that:

$$S'(x,t) = \sum_{m=1}^{r} p_m(x) \cdot \alpha_m(t) + \epsilon_m(x,t)$$

with r the number of dimensions of the EOF ($r \leq min(n,p)$), $p_m(x)$ the spatial term of EOF and $\alpha_m(t)$ the temporal term of EOF for dimension m.
$\epsilon_m(x,t)$ is an error term.

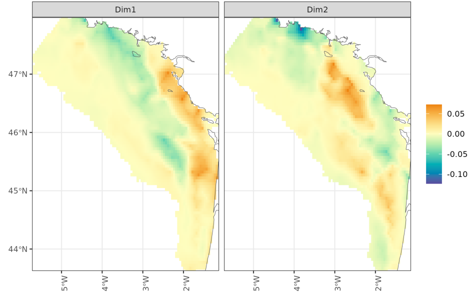
Constraints:

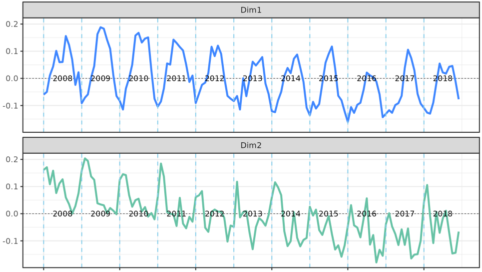
- minimize $E = \sum_m \sum_x \sum_y \epsilon_m(x,t)$
- spatial terms and temporal terms are orthogonal

$$\langle p_i(\cdot); p_j(\cdot) \rangle = 0 \quad i \neq j$$

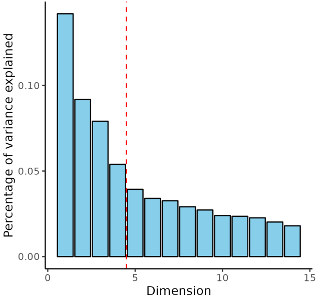
$$\langle \alpha_i(\cdot); \alpha_j(\cdot) \rangle = 0 \quad i \neq j$$

Illustration

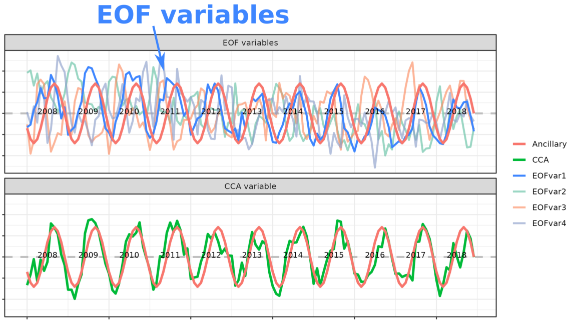




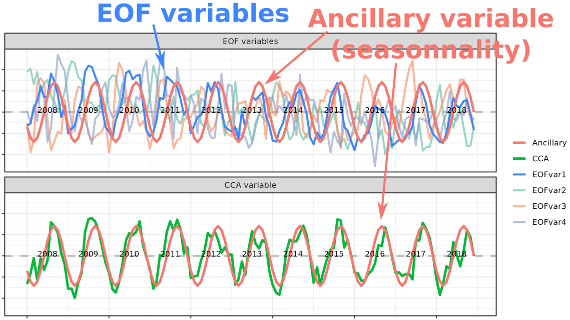
(Top) Spatial factors for the two first dimensions of the EOF. (Bottom) Loadings for the two first dimensions of the EOF. Blue dashed vertical lines corresponds to the month of January for each year.



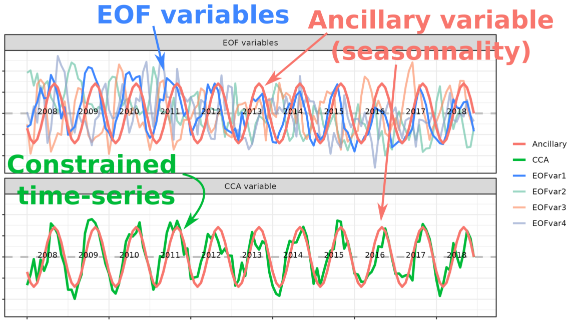
Constraining the EOF with an ancillary variable



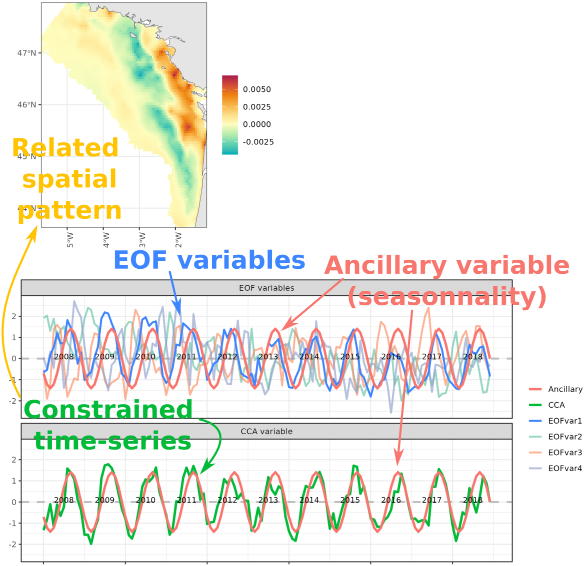
Constraining the EOF with an ancillary variable



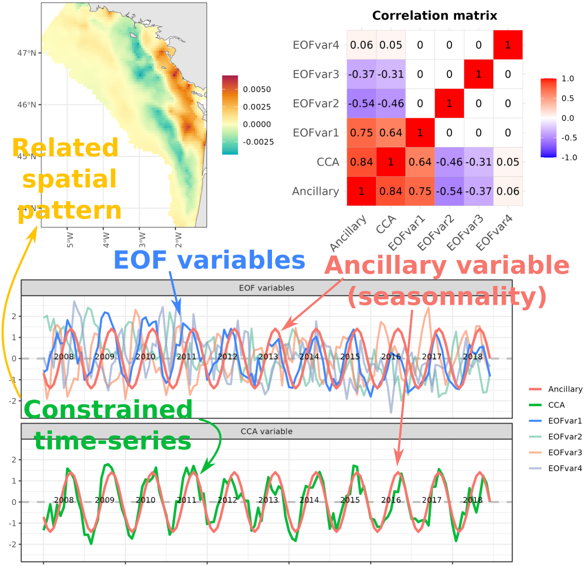
Constraining the EOF with an ancillary variable



Constraining the EOF with an ancillary variable



Constraining the EOF with an ancillary variable



Conclusion

The ancillary variable could be :

- Environmental covariate time-series
- Some oceanographic process (the cold pool extent, upwelling)
- Any idea?

Many other things that can be done through such technics !

- Multivariate EOF
- Trend EOF
- Non-linear EOF
- Complex EOF (propagating patterns)
- Stronger orthogonality constraints (EOM)
- What's next ? Any idea ?

Conclusion

The ancillary variable could be :

- Environmental covariate time-series
- Some oceanographic process (the cold pool extent, upwelling)
- Any idea?

Many other things that can be done through such technics !

- Multivariate EOF
- Trend EOF
- Non-linear EOF
- Complex EOF (propagating patterns)
- Stronger orthogonality constraints (EOM)
- What's next ? Any idea ?

Thank you for your attention!

Bibliography

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., . . . & Vuichard, N. (2013). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Climate dynamics, 40, 2123-2165.

Garrigues, S., Allard, D., & Baret, F. (2008). Modeling temporal changes in surface spatial heterogeneity over an agricultural site. Remote Sensing of Environment, 112(2), 588-602.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. Biological conservation, 142(10), 2282-2292.

Thorson, J. T., Cheng, W., Hermann, A. J., Ianelli, J. N., Litzow, M. A., O'Leary, C. A., & Thompson, G. G. (2020). Empirical orthogonal function regression: Linking population biology to spatial varying environmental conditions using climate projections. Global Change Biology, 26(8), 4638-4649.

Wikle, Christopher K., Andrew Zammit-Mangion, and Noel Cressie. Spatio-temporal statistics with R. Chapman and Hall/CRC, 2019.

Some ideas of the maths behind this

Aim: "create new variables that are linear combinations of two (multivariate) data sets such that the correlations between these new variables are maximized" (Wickle et al., 2019).

Let's consider two (spatio-)temporal variables $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$.
Now consider two new variables that are combinations of $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$

$$a_k(t_j) = \sum_{i=1}^{n} \xi_{ik} S^{(1)}(x_i; t_j) = \boldsymbol{\xi}'_k \mathbf{s}^{(1)}_{t_j}$$

$$b_k(t_j) = \sum_{\ell=1}^{m} \psi_{\ell k} S^{(2)}(r_\ell; t_j) = \boldsymbol{\psi}'_k \mathbf{s}^{(2)}_{t_j}$$

The weights (*i.e.* the $\text{k}^t h$ canonical correlation) are the correlation between $a_k(.)$ and $b_k(.)$ with $k \in \{1, \cdots, \min\{n, m\}\}$:

$$r_k = \text{corr}(\mathbf{a}_k, \mathbf{b}_k) = \frac{\text{cov}(\mathbf{a}_k, \mathbf{b}_k)}{\sqrt{\text{var}(\mathbf{a}_k)}\sqrt{\text{var}(\mathbf{b}_k)}}$$

Some ideas of the maths behind this

<u>**Aim:**</u> "create new variables that are linear combinations of two (multivariate) data sets such that the correlations between these new variables are maximized" (Wickle et al., 2019).

Let's consider two (spatio-)temporal variables $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$.
Now consider two new variables that are combinations of $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$

$$a_k(t_j) = \sum_{i=1}^{n} \xi_{ik} S^{(1)}(x_i;t_j) = \boldsymbol{\xi}'_k \mathbf{s}^{(1)}_{t_j}$$

$$b_k(t_j) = \sum_{\ell=1}^{m} \psi_{\ell k} S^{(2)}(r_\ell;t_j) = \boldsymbol{\psi}'_k \mathbf{s}^{(2)}_{t_j}$$

The weights (*i.e.* the $\mathrm{k}^t h$ canonical correlation) are the correlation between $a_k(.)$ and $b_k(.)$ with $k \in \{1,\cdots,\min\{n,m\}\}$:

$$r_k = \mathrm{corr}(\mathbf{a}_k,\mathbf{b}_k) = \frac{\mathrm{cov}(\mathbf{a}_k,\mathbf{b}_k)}{\sqrt{\mathrm{var}(\mathbf{a}_k)}\sqrt{\mathrm{var}(\mathbf{b}_k)}}$$

Some ideas of the maths behind this

Aim: "create new variables that are linear combinations of two (multivariate) data sets such that the correlations between these new variables are maximized" (Wickle et al., 2019).

Let's consider two (spatio-)temporal variables $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$.
Now consider two new variables that are combinations of $S^{(1)}(x,t)$ and $S^{(2)}(x,t)$

$$a_k(t_j) = \sum_{i=1}^{n} \xi_{ik} S^{(1)}(x_i;t_j) = \boldsymbol{\xi}'_k \mathbf{s}^{(1)}_{t_j}$$

$$b_k(t_j) = \sum_{\ell=1}^{m} \psi_{\ell k} S^{(2)}(r_\ell;t_j) = \boldsymbol{\psi}'_k \mathbf{s}^{(2)}_{t_j}$$

The weights (*i.e.* the k$^t h$ canonical correlation) are the correlation between $a_k(.)$ and $b_k(.)$ with $k \in \{1, \cdots, \min\{n,m\}\}$:

$$r_k = \mathrm{corr}(\mathbf{a}_k, \mathbf{b}_k) = \frac{\mathrm{cov}(\mathbf{a}_k, \mathbf{b}_k)}{\sqrt{\mathrm{var}(\mathbf{a}_k)}\sqrt{\mathrm{var}(\mathbf{b}_k)}}$$

Some ideas of the maths behind this

Aim: "create new variables that are linear combinations of two (multivariate) data sets such that the correlations between these new variables are maximized" (Wickle et al., 2019).

Let's consider two (spatio-)temporal variables $S^{(1)}(x, t)$ and $S^{(2)}(x, t)$.
Now consider two new variables that are combinations of $S^{(1)}(x, t)$ and $S^{(2)}(x, t)$

$$a_k(t_j) = \sum_{i=1}^{n} \xi_{ik} S^{(1)}(x_i; t_j) = \boldsymbol{\xi}_k' \mathbf{s}_{t_j}^{(1)}$$

$$b_k(t_j) = \sum_{\ell=1}^{m} \psi_{\ell k} S^{(2)}(r_\ell; t_j) = \boldsymbol{\psi}_k' \mathbf{s}_{t_j}^{(2)}$$

The weights (*i.e.* the k$^t h$ canonical correlation) are the correlation between $a_k(.)$ and $b_k(.)$ with $k \in \{1, \cdots, \min\{n, m\}\}$:

$$r_k = \text{corr}(\mathbf{a}_k, \mathbf{b}_k) = \frac{\text{cov}(\mathbf{a}_k, \mathbf{b}_k)}{\sqrt{\text{var}(\mathbf{a}_k)}\sqrt{\text{var}(\mathbf{b}_k)}}$$