Empirical Orthogonal Functions and their latest developments

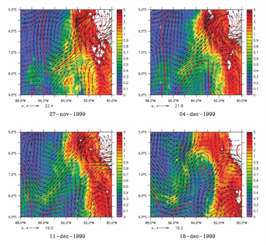
Dealing with orthogonality and sparse data

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

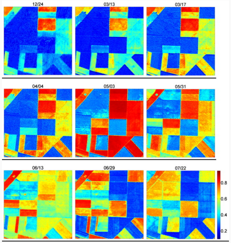
University Bretagne Sud, Lab-STICC, Vannes.
University Bretagne Sud, LMBA, Vannes.
BioSP, INRAE, Avigon, France.
Alaska Fisheries Science Center, NOAA, Seattle, USA.

Context

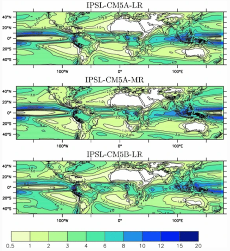
Oceanography



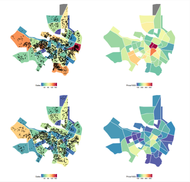
Agronomy



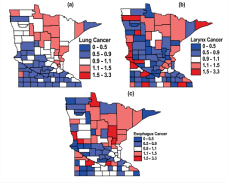
Climatology



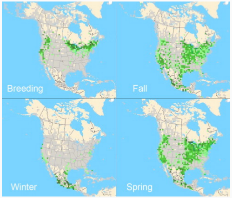
Economics



Epidemiology

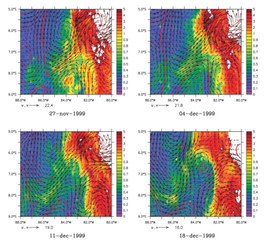


Ecology

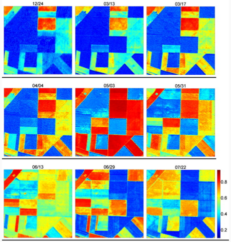


Context

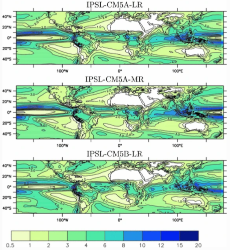
Oceanography



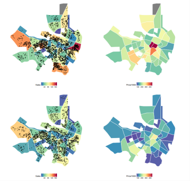
Agronomy



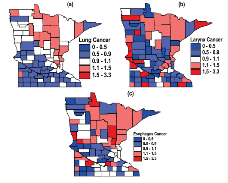
Climatology



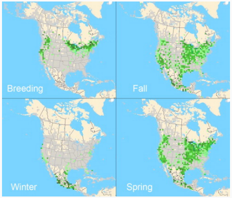
Economics



Epidemiology

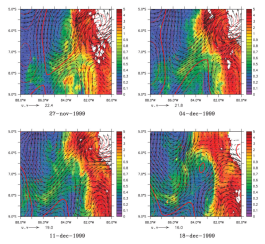


Ecology

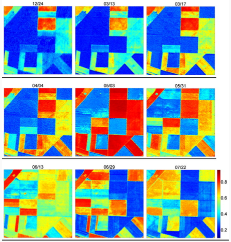


Context

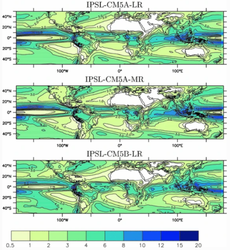
Oceanography



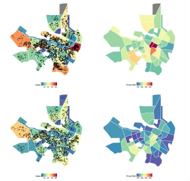
Agronomy



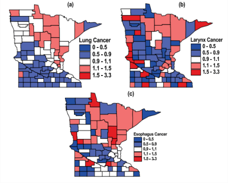
Climatology



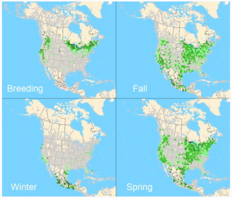
Economics



Epidemiology



Ecology



Context

How to best represent and summarise these data ?

How to best represent and summarise these data ?

Empirical Orthogonal Functions (EOFs) is the keystone method for dimension reduction of spatio-temporal data.

First introduced by **Lorenz** in the 50's [Lorenz, 1956].

Extensive literature has then focused on this method either to apply it on climate applications, either to extend this method [Hannachi et al., 2007].

Context

How to best represent and summarise these data ?

Empirical Orthogonal Functions (EOFs) is the keystone method for dimension reduction of spatio-temporal data.

First introduced by **Lorenz** in the 50's [Lorenz, 1956].

Extensive literature has then focused on this method either to apply it on climate applications, either to extend this method [Hannachi et al., 2007].

Few extensions:

- Extended and multivariate EOFs [Fraedrich et al., 1997, Sparnocchia et al., 2003]
- Complex/hilbert EOFs [Bürger, 2021]
- Regularized EOFs [Hannachi, 2016]
- Non-linear EOFs [Bueso et al., 2020]

How to best represent and summarise these data ?

Empirical Orthogonal Functions (EOFs) is the keystone method for dimension reduction of spatio-temporal data.

First introduced by **Lorenz** in the 50's [Lorenz, 1956].

Extensive literature has then focused on this method either to apply it on climate applications, either to extend this method [Hannachi et al., 2007].

Few extensions:

- Extended and multivariate EOFs [Fraedrich et al., 1997, Sparnocchia et al., 2003]
- Complex/hilbert EOFs [Bürger, 2021]
- Regularized EOFs [Hannachi, 2016]
- Non-linear EOFs [Bueso et al., 2020]

New methods from ecology

Empirical Orthogonal Functions: the basics

Let's denote a spatio-temporal process $S = (S(x,t); x \in \mathbb{R}^2,$
$t \in \{t_1, \cdots, t_p\})$

The temporal average of S in denoted:

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

The time-centered space-time field:

$$\mathbf{S}' = \left(\mathbf{s}_{t_1} - \overline{\mathbf{s}}^t, ..., \mathbf{s}_{t_p} - \overline{\mathbf{s}}^t\right).$$

Then, $\mathbf{S}'$ has the form:

$$\mathbf{S}' = \begin{pmatrix} S'(x_1,t_1) & S'(x_1,t_2) & \cdots & S'(x_1,t_p) \\ S'(x_2,t_1) & S'(x_2,t_2) & \cdots & S'(x_2,t_p) \\ \vdots & \ddots & \ddots & \vdots \\ S'(x_n,t_1) & S'(x_n,t_2) & \cdots & S'(x_n,t_p) \end{pmatrix}$$

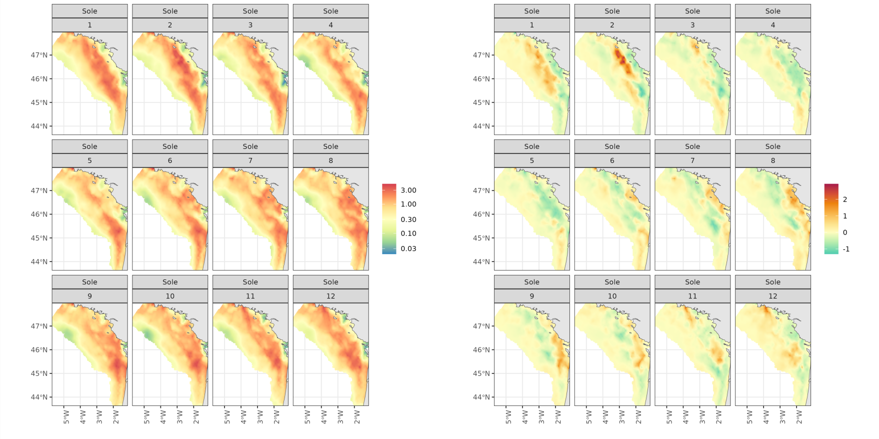


Figure 1: (left) Montly spatial log-predictions $\log S(x, t)$ of the hierarchical model. (right) Monthly anomalies of the spatial predictions $S^*(x, t)$. Each panel corresponds to the average distribution of prediction of anomalies for a month over the period 2008 - 2018.

The spatio-temporal field is decomposed so that:

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

where $\mathbf{u}_m$ capture the variance of $\mathbf{S}'$ and $\langle u_i(\cdot); u_j(\cdot)\rangle = 0$ and $\langle v_i(\cdot); v_j(\cdot)\rangle$ with $i \neq j$

This falls back to a diagonalisation through singular value decomposition:

$$\mathbf{S}' = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \quad \text{(SVD)}$$

- $\mathbf{U}_{(n\times r)}$ contains the spatial factors,
- $\boldsymbol{\Sigma}_{(r\times r)}$ contains the singular values of $\mathbf{S}'$.
- $\mathbf{V}_{(p\times r)}$ contains the temporal loadings

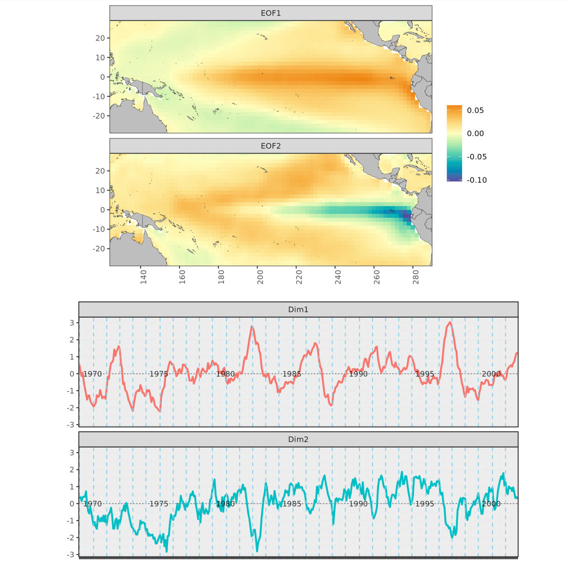


Figure 2: SST in the South Pacific. Results of the EOFs (Top) Spatial factors of two first dimensions. (Bottom) Related temporal loadings. Blue dashed vertical lines corresponds to the month of January for each year.

SST South Pacific

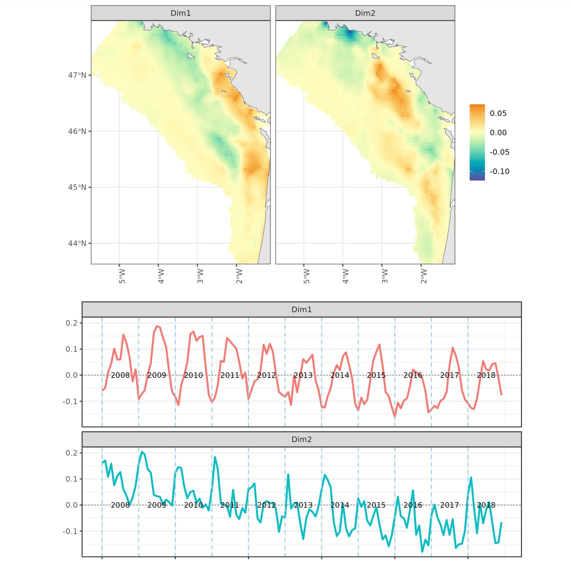


Figure 2: Sole in the Bay of Biscay. Results of the EOFs (Top) Spatial factors of two first dimensions. (Bottom) Related temporal loadings. Blue dashed vertical lines corresponds to the month of January for each year.

Sole
Bay of Biscay

Setting spatial orthogonal constraints on EOFs

Setting spatial orthogonal constraints on EOFs

EOFs are statistically orthogonal *i.e.* for $i \neq j, \left\{\mathbf{u}_i, \mathbf{u}_j\right\} = 0$

Bu they are not spatially orthogonal *i.e.* the cross-covariance of the different EOFs maps is not necessarily 0.

$$\gamma_{ij}(r) = \frac{1}{2} E\left[\mathbf{u}_i(x) - \mathbf{u}_i(x+r)\right]\left[\mathbf{u}_j(x) - \mathbf{u}_j(x+r)\right]$$

EOM consist in a two step analysis:

- perform an EOFs on $\mathbf{S}'$ $\rightarrow$ obtain statistically decorrelated planes
- compute the covariogram and the cross-covariogram of $\mathbf{U}$ for a specific ditance r that we denote $\boldsymbol{\Gamma}_r$ of dimension $(p \times p)$.

$$\boldsymbol{\Gamma}_r = \boldsymbol{U}_{\mathrm{eom}} \boldsymbol{\Sigma}_{eom} \mathbf{V}_{\mathrm{eom}}^T$$

Planes are then ordered by increasing variance explained in $\mathbf{S}'$

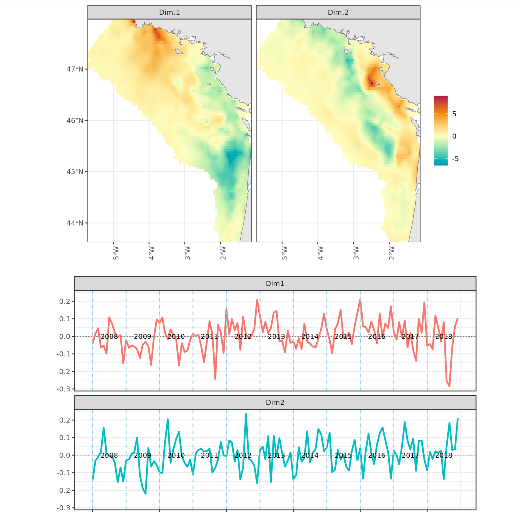


Figure 3: Sole in the Bay of Biscay. (Top) Spatial factors obtained by EOM for the two first dimensions. (Bottom) Temporal loadings for the first two dimensions.

Sole

Bay of Biscay

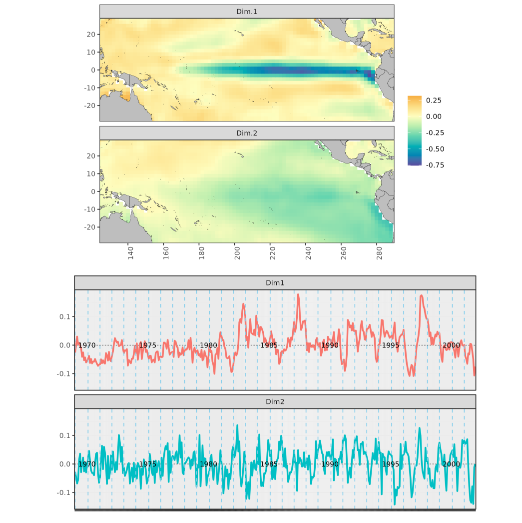


Figure 3: SST in the South Pacific. (Top) Spatial factors obtained by EOM for the two first dimensions. (Bottom) Temporal loadings for the first two dimensions.

SST

South Pacific

Inferring EOFs from sparse data

Inferring EOFs from sparse data

In many cases, EOFs must be computed from sparse data ('VAST', Jim Thorson).

In many cases, EOFs must be computed from sparse data (‘VAST‘, Jim Thorson).

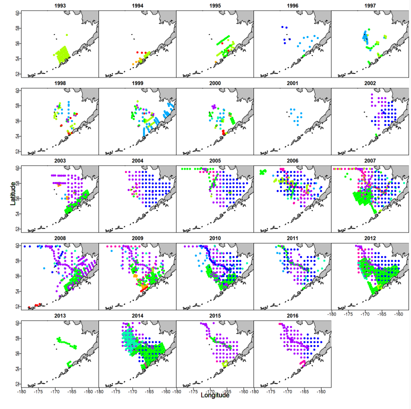


Figure 4: (Top) Location of survey tows for nine seasons (colors) in each year (panel) in the Bering Sea case study [Thorson et al., 2020].

In many cases, EOFs must be computed from sparse data (‘VAST‘, Jim Thorson).

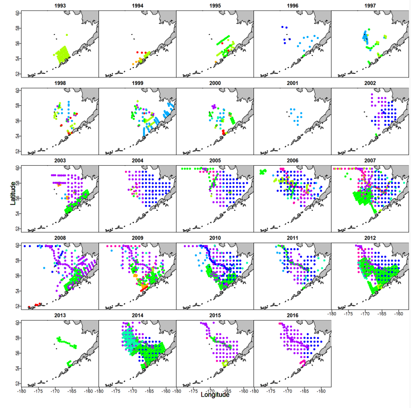


Figure 4: (Top) Location of survey tows for nine seasons (colors) in each year (panel) in the Bering Sea case study [Thorson et al., 2020].

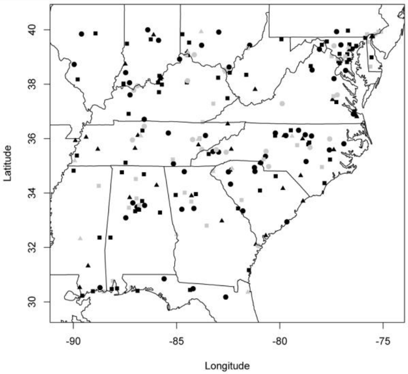


Figure 5: Sites reporting concentration of ozone (dots and squares) and PM2.5 (triangles and squares) used in our case study [Berrocal et al., 2010].

Let's define observations Y:

$$Y|S, \boldsymbol{\theta} \sim \mathcal{L}_Y(S, \boldsymbol{\theta}_{obs})$$

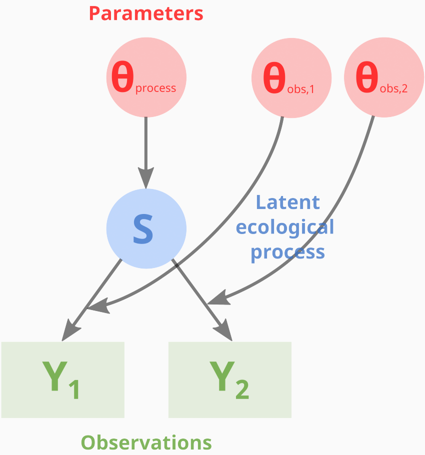
a latent field S:

$$S|\boldsymbol{\theta}_{process} \sim \mathcal{L}_S(\boldsymbol{\theta}_{process})$$

and parameters:

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_{obs}, \boldsymbol{\theta}_{process})$$

There can be several data sources Y_1 and Y_2, in which case each has its own probability distribution ($\mathcal{L}_{Y_1}$,$\mathcal{L}_{Y_2}$) and observation parameters ($\boldsymbol{\theta}_{obs,1}$, $\boldsymbol{\theta}_{obs,2}$).



We consider a **hierarchical model** where the **latent field** take the form:

$$log(S(x,t)) = \beta(t) + \omega(x) + \sum_{f=1}^{F} \lambda(t,f)\epsilon(x,f)$$

where f are the dimension of the EOFs with F dimensions, $\boldsymbol{\omega} \sim \mathcal{MG}(0, \boldsymbol{\Sigma}_\omega)$ and $\boldsymbol{\epsilon} \sim \mathcal{MG}(0, \boldsymbol{\Sigma}_\epsilon)$.

$\lambda(t,f)$ are the loadings and $\epsilon(x,g)$ are the EOF.

---

and **the observations** are zero-inflated and takes the form:

$$\Pr(Y = y_i) = \begin{cases} 1 - p & \text{if } y_i = 0 \\ p \times \mathcal{L}(y_i; \log S(x_i, t_i); \sigma^2) & \text{if } y_i > 0 \end{cases}$$

Where Y is an observations as random variable and y_i is the realized observation. p is the probability to obtain a positive observation. $\mathcal{L}$ is the probability of the positive observations.

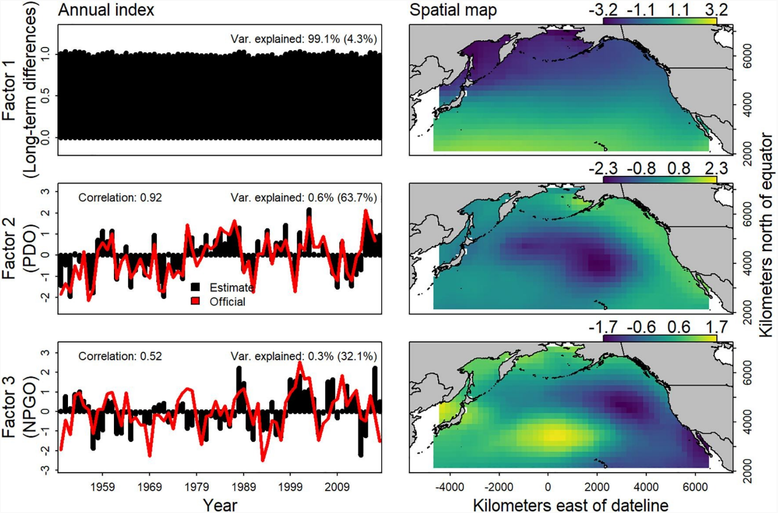


Figure 6: EOFs for sea surface temperature in the North Pacific (rows), where each axis includes the loadings (left columns) as well as a spatial factors (right column) [Thorson et al., 2021]. PDO: Pacific Decadal Oscillation. NPGO: North Pacific Gyre Oscillation.

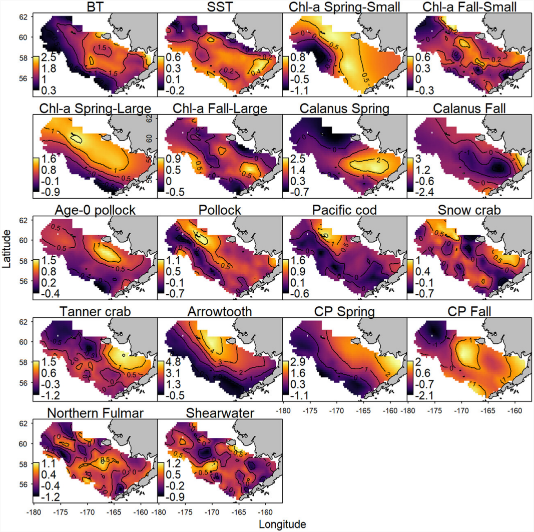


Figure 7: Spatial factors for each variable of the primary mode of ecosystem variability [Thorson et al., 2021].

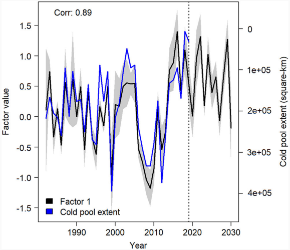
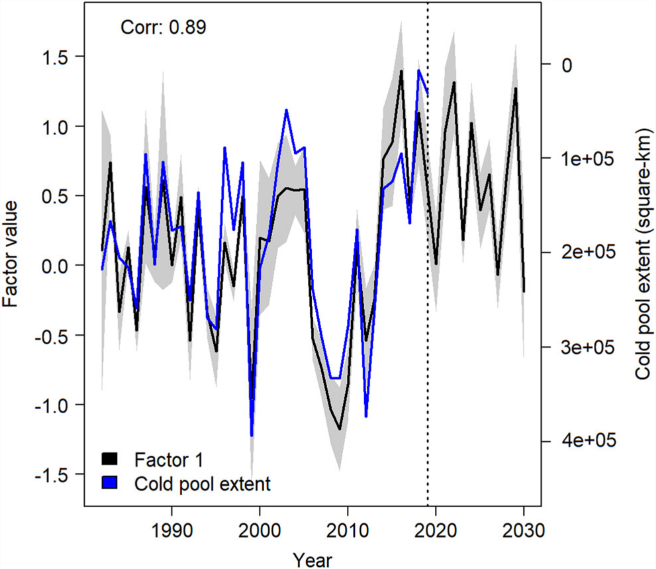


Figure 8: Dominant mode of ecosystem variability for spatio-temporal variation (black lines) and cold-pool extent (blue line) [Thorson et al., 2021].

Illustration: ecological data in the Bering sea.



Take home message

- **EOFs have been extended** in several ways in the past decade.

Take home message

- **EOFs have been extended** in several ways in the past decade.
- EOM provide **spatially decorrelated patterns** i.e. they do not share common information.

- **EOFs have been extended** in several ways in the past decade.
- EOM provide **spatially decorrelated patterns** i.e. they do not share common information.

  ➠ Still very **sensitive to the spatial lag** and it is sometimes hard to interpret the spatial patterns.

Take home message

- **EOFs have been extended** in several ways in the past decade.
- EOM provide **spatially decorrelated patterns** i.e. they do not share common information.

  ➠ Still very **sensitive to the spatial lag** and it is sometimes hard to interpret the spatial patterns.
- EOFs within a **hierarchical statistical framework** allows for both mapping of a continuous field from sparse data and decomposition in EOFs.

Take home message

- **EOFs have been extended** in several ways in the past decade.
- EOM provide **spatially decorrelated patterns** i.e. they do not share common information.

  ➠ Still very **sensitive to the spatial lag** and it is sometimes hard to interpret the spatial patterns.
- EOFs within a **hierarchical statistical framework** allows for both mapping of a continuous field from sparse data and decomposition in EOFs.

  ▲ The patterns are **not orthogonal**, this requires additional rotation.

Take home message

- **EOFs have been extended** in several ways in the past decade.
- EOM provide **spatially decorrelated patterns** i.e. they do not share common information.

  ➠ Still very **sensitive to the spatial lag** and it is sometimes hard to interpret the spatial patterns.
- EOFs within a **hierarchical statistical framework** allows for both mapping of a continuous field from sparse data and decomposition in EOFs.

  ▲ The patterns are **not orthogonal**, this requires additional rotation.
- Further research avenues: perform EOFs that **accounts for both spatial and temporal correlation**.

Thank you for your attention!

Any questions?

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010).
A bivariate space-time downscaler under space and time misalignment.
The annals of applied statistics, 4(4):1942.

Bueso, D., Piles, M., and Camps-Valls, G. (2020).
Nonlinear pca for spatio-temporal analysis of earth observation data.
IEEE Transactions on Geoscience and Remote Sensing, 58(8):5752–5763.

Bürger, G. (2021).
Intraseasonal oscillation indices from complex eofs.
Journal of Climate, 34(1):107–122.

References II

- Fraedrich, K., McBride, J. L., Frank, W. M., and Wang, R. (1997).
  Extended eof analysis of tropical disturbances: Toga coare.
  Journal of the atmospheric sciences, 54(19):2363–2372.
- Hannachi, A. (2016).
  Regularised empirical orthogonal functions.
  Tellus A: Dynamic Meteorology and Oceanography, 68(1):31723.
- Hannachi, A., Jolliffe, I. T., Stephenson, D. B., et al. (2007).
  Empirical orthogonal functions and related techniques in atmospheric science: A review.
  International journal of climatology, 27(9):1119–1152.

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

Sparnocchia, S., Pinardi, N., and Demirov, E. (2003).
Multivariate empirical orthogonal function analysis of the upper thermocline structure of the mediterranean sea from observations and model simulations.
In *Annales Geophysicae*, volume 21, pages 167–187. Copernicus GmbH.

Thorson, J. T., Adams, C. F., Brooks, E. N., Eisner, L. B., Kimmel, D. G., Legault, C. M., Rogers, L. A., and Yasumiishi, E. M. (2020).
Seasonal and interannual variation in spatio-temporal models for index standardization and phenology studies.
ICES Journal of Marine Science, 77(5):1879–1892.

Thorson, J. T., Arimitsu, M. L., Barnett, L. A., Cheng, W., Eisner, L. B., Haynie, A. C., Hermann, A. J., Holsman, K., Kimmel, D. G., Lomas, M. W., et al. (2021).
Forecasting community reassembly using climate-linked spatio-temporal ecosystem models.
Ecography, 44(4):612–625.