

## Supplementary Material 2

### EOFs interpretation

It is important to note that the signs of the loadings and factors are arbitrary, meaning that switching the sign of both  $\mathbf{v}_j$  and  $\mathbf{u}_j$  does not change the overall model. This label (or sign) indeterminacy is common in factor models and should be kept in mind when interpreting the results, as the sign of the factors may flip between different analyses or software implementations.

Also, in our case the singular values are not related to the variance explained by each dimension. This is a consequence of the centering that is realized on rows rather than on columns. The latter is the classical centering realized in PCA. In this specific case,  $\lambda_m^2 = \sigma_m^2$  where  $\sigma_m^2$  is the variance explained by the  $m^{\text{th}}$  dimension of the EOFs.

Still, computing the variance captured by each dimension is possible by back-transforming the factor  $\mathbf{u}_m$  in the space of the anomalies as described by Bez, Renard, and Ahmed-Babou (2023):

$$\hat{\mathbf{S}}'_{(m)} = \lambda_m \mathbf{u}_m \mathbf{v}_m^T$$

where  $\hat{\mathbf{S}}'_{(m)}$  is the reconstructed spatio-temporal anomaly matrix based on the  $m^{\text{th}}$  dimension of the EOFs

Then, the percentage of variance explained by this dimension is defined as:

$$p_i = \frac{\text{tr} \left( C_{\hat{S}'_{(m)}} \right)}{\text{tr} \left( C'_{\mathbf{S}} \right)}$$

where  $\text{tr}()$  is the trace operator,  $C_{\mathbf{S}'}$  is the covariance of  $\mathbf{S}'$ , and  $C_{\hat{S}'_{(m)}}$  is the covariance of the reconstructed spatio-temporal anomaly matrix based on the  $m^{\text{th}}$  dimension of the EOFs. This is the method used to compute the variance related to each EOFs dimension throughout the paper.

An important consideration in EOFs analysis is the number of dimensions retained for interpretation. Several rules of thumb can guide this decision. Many are based on the proportion of variance captured by each dimension. This is often represented through a scree plot. One approach is to keep the dimensions that capture the variance of more than what a single variable would capture if all the variables were independent. This means retaining dimensions that capture more than one out of the number of variables. Another method involves examining the variance graph and selecting dimensions up to the point where there is a noticeable drop in variance explained, indicating a significant change in the slope of the scree plot. Also, one may choose to retain only the dimensions that are interpretable from an ecological point of view. Finally, it is possible to randomly perturb the values for each spatial location at each time step, and compare the scree plot

with the original data (Hastie, 2009). These criteria are not mutually exclusive. For simplicity, we represent only the first two dimensions, with additional dimensions being provided in the Supplementary Material.

Note that EOFs should be interpreted as patterns that capture variance rather than they capture ecological or physical processes i.e. each individual EOFs dimension may not necessarily represent a physically or dynamically relevant spatial pattern (Monahan et al., 2009; Roundy, 2015). For example, a single EOFs dimension may not correspond directly to a specific physical phenomenon; instead, such process might be captured by a linear combination of multiple EOFs. Additionally, there is no constraint separating long-term and short-term signals, so both could be combined within a single dimension. Identifying individual EOFs with underlying processes should be done with great care, always keeping in mind that EOFs capture variance rather than they represent an ecological process per se.

An example of EOFs interpretation in climate science is the North Atlantic Oscillation (NAO) index (Hurrell and Deser, 2010). The NAO is a weather phenomenon in the North Atlantic Ocean characterized by fluctuations in the atmospheric pressure difference at sea level between the Icelandic Low and the Azores High. The NAO has been shown to influence ecological dynamics in both marine and terrestrial systems (Ottersen et al., 2001). The first EOFs dimension of the sea-level pressure field over the North Atlantic

typically represents the NAO pattern (Figure 8 in Saeed, Kucharski, and Almazroui, 2023). When the loadings associated with this dimension are positive, it indicates a strong Azores high and a deep Icelandic low, corresponding to the positive phase of the NAO. Conversely, negative loadings indicate a negative phase of the NAO, characterized by a deep Azores low and a strong Icelandic high. This EOFs dimension captures the primary mode of variability in the North Atlantic sea-level pressure field and is crucial for understanding and predicting regional climate variations.

In standard multivariate statistics, PCA is often combined with clustering analysis (*e.g.*, k-means, Hierarchical Clustering on Principal Components - HCPC) to group individuals into homogeneous clusters. A similar approach can be applied to EOFs to either group locations with similar temporal trends or group time steps with similar spatial patterns (Lindegren et al., 2022; Alglave et al., 2024). This can be particularly useful for identifying functional habitats from species spatio-temporal distributions (Alglave et al., 2024).

## References

- Alglave, Baptiste et al. (2024). "Investigating fish reproduction phenology and essential habitats by identifying the main spatio-temporal patterns of fish distribution". In: *ICES Journal of Marine Science*, fsae099.
- Bez, Nicolas, Didier Renard, and Dedah Ahmed-Babou (2023). "Empirical Orthogonal Maps (EOM) and Principal Spatial Patterns: Illustration for Octopus Distribution Off Mauritania Over the Period 1987–2017". In: *Mathematical Geosciences* 55.1, pp. 113–128.
- Hastie, Trevor (2009). *The elements of statistical learning: data mining, inference, and prediction*.
- Hurrell, James W and Clara Deser (2010). "North Atlantic climate variability: the role of the North Atlantic Oscillation". In: *Journal of marine systems* 79.3-4, pp. 231–244.

- Lindegren, Martin et al. (2022). “A spatial statistical approach for identifying population structuring of marine fish species: European sprat as a case study”. In: ICES Journal of Marine Science 79.2, pp. 423–434.
- Monahan, A. H. et al. (2009). “Empirical Orthogonal Functions: The Medium is the Message”. In: Journal of Climate 22.24, pp. 6501–6514. doi: 10.1175/2009JCLI3062.1. URL: <https://journals.ametsoc.org/view/journals/clim/22/24/2009jcli3062.1.xml>.
- Ottersen, Geir et al. (2001). “Ecological effects of the North Atlantic oscillation”. In: Oecologia 128, pp. 1–14.
- Roundy, P. E. (2015). “On the Interpretation of EOF Analysis of ENSO, Atmospheric Kelvin Waves, and the MJO”. In: Journal of Climate 28.3, pp. 1148–1165. doi: 10.1175/JCLI-D-14-00398.1. URL: <https://journals.ametsoc.org/view/journals/clim/28/3/jcli-d-14-00398.1.xml>.
- Saeed, Sajjad, Fred Kucharski, and Mansour Almazroui (2023). “Impacts of mid-latitude circulation on winter temperature variability in the Arabian Peninsula: the explicit role of NAO”. In: Climate Dynamics 60.1, pp. 147–164.