Let's generalise these "results"

Heuristic Derivation of EOF or Principal Component Analysis

Assume we have M time series $u_i(t) = u_i(t, x_i)$

We define an EOF as a set of functions that allow the following transformation

$$u_i(t, x_i) = \sum_{n=1}^{M} A_n(t) \cdot \phi_n(x_i)$$

vector space of data — vector space of statistical "modes"

where

- n denotes the n-th EOF mode
- $A_n(t)$ denotes the amplitude of the n-th EOF mode (eigenfunction)
- $\phi_n(x_i)$ is the pattern (eigenvector) of the n-th EOF mode
- i denotes the ~~variable~~ i-th variable

The functions $A_n(t)$ and $\phi_n(x_i)$ are subject to the following orthogonality conditions

(1.) $$\sum_{i=1}^{M} \phi_n(x_i)\, \phi_m(x_i) = \delta_{nm} = \begin{cases} 0 & n \neq m \\ 1 & n = m \end{cases}$$

(2.) $$\frac{1}{K} \sum_{k=1}^{K} A_n(t_k)\, A_m(t_k) = \delta_{nm} \frac{1}{K} \sum_{k=1}^{K} A_n^2(t_k)$$

$$\langle A_n(t)\, A_m(t) \rangle = \delta_{nm}\, \lambda_n$$

I will show next that the above definitions and constraints constitute an eigenvalue problem for the cross-correlation matrix.

Consider time series at two locations, i.e., $u_i(t, x_i)$ and $u_j(t, x_j)$. The product of these data are

$$u_i(t, x_i) \cdot u_j(t, x_j) = \sum_{n=1}^{M} \sum_{m=1}^{M} A_n(t) \phi_n(x_i) A_m(t) \phi_m(x_j)$$

taking the time (or ensemble) average, i.e., $\langle \cdot \rangle \equiv \frac{1}{K} \sum_{k=1}^{K} \cdot (t_k)$

$$\langle u_i(t, x_i) \cdot u_j(t, x_j) \rangle = \sum_{n=1}^{M} \sum_{m=1}^{M} \langle A_n(t) A_m(t) \rangle \phi_n(x_i) \phi_m(x_j)$$

using (2.)

$$= \sum_{n=1}^{M} \sum_{m=1}^{M} \delta_{nm} \lambda_n \phi_n(x_i) \phi_m(x_j)$$

$$= \sum_{n=1}^{M} \lambda_n \phi_n(x_i) \phi_n(x_j)$$

For an arbitrary mode m at location j we have

$$\phi_m(x_j) \langle u_i u_j \rangle = \phi_m(x_j) \sum_{n=1}^{M} \lambda_n \phi_n(x_i) \phi_n(x_j) \qquad m = 1, 2, \ldots, M$$

$$= \sum_{n=1}^{M} \lambda_n \phi_n(x_i) \phi_n(x_j) \phi_m(x_j)$$

start class #21

And summing over all locations j

$$\sum_{j=1}^{M} \phi_m(x_j) \langle u_i u_j \rangle = \sum_{j=1}^{M} \sum_{n=1}^{M} \lambda_n \phi_n(x_i) \phi_n(x_j) \phi_m(x_j) \qquad m = 1, 2, \dots M$$

$$= \sum_{n=1}^{M} \lambda_n \phi_n(x_i) \sum_{j=1}^{M} \phi_n(x_j) \phi_m(x_j)$$

using (1.)

$$= \sum_{n=1}^{M} \lambda_n \phi_n(x_i) \, \delta_{nm}$$

end #20

$$\boxed{\sum_{j=1}^{M} \phi_m(x_j) \langle u_i u_j \rangle = \lambda_m \phi_m(x_i) \qquad m = 1, 2, \dots M}$$

Which defines an eigenvalue problem

M is the # of time series

$$\begin{array}{c} \\ m=1=i \\ m=2=i \\ \\ \\ \end{array}
\begin{array}{c} \quad i \quad j \\ \begin{pmatrix} \langle u_1 u_1 \rangle & \langle u_1 u_2 \rangle & \dots & \langle u_{1M} \rangle \\ \langle u_2 u_1 \rangle & \langle u_2 u_2 \rangle & \dots & \langle u_{2M} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle u_M u_1 \rangle & \langle u_M u_2 \rangle & \dots & \langle u_M u_M \rangle \end{pmatrix} \end{array}
\begin{pmatrix} \phi_m(x_1) \\ \phi_m(x_2) \\ \vdots \\ \phi_m(x_M) \end{pmatrix} = \lambda_m$$

$$= \begin{pmatrix} \lambda_m & 0 & \dots & 0 \\ 0 & \lambda_m & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & \lambda_m \end{pmatrix} \begin{pmatrix} \phi_m(x_1) \\ \phi_m(x_2) \\ \vdots \\ \phi_m(x_M) \end{pmatrix} = \lambda_m \begin{pmatrix} \phi_m(x_1) \\ \phi_m(x_2) \\ \vdots \\ \phi_m(x_M) \end{pmatrix}$$

$m = 1, 2, 3 \dots M$

The set of functions $\{ \phi_1, \phi_2, \dots, \phi_M \}$ form the basis vectors of a "new" vector space such that the data can be represented as

$$u_i(x_i, t_\ell) = \sum_{n=1}^{M} A_n(t_\ell)\, \phi_n(x_i)$$

where

$$A_n(t_\ell) = \sum_{i=1}^{M} u_i(x_i, t_\ell)\, \phi_n(x_i)$$

We still need to find the eigenfunctions ~~ϕ_n~~ or eigenvectors $\phi_n(x_i)$

e.g. $\longrightarrow$ singular value decomposition (better use TRED2 + TQLI of Numerical Recipes)

Let us first, however, have a close look at the matrix

$$R_{ij} = \langle u_i u_j \rangle = \langle u_i(t, x_i)\, u_j(t, x_j) \rangle = \frac{1}{K} \sum_{\ell=1}^{K} u(t_k, x_i)\, u(t_\ell, x_j)$$

The diagonal elements R_{ii} ~~are~~ represent the variance of variable i:

$$R_{ii} = \langle u_i u_i \rangle = \left\langle \sum_{n=1}^{M} A_n \phi_n(x_i) \cdot \sum_{n=1}^{M} A_n \phi_n(x_i) \right\rangle$$

$$= \left\langle \sum_{n=1}^{M} \sum_{m=1}^{M} A_n \phi_n(x_i)\, A_m \phi_m(x_i) \right\rangle$$

$$= \sum_{n=1}^{M} \sum_{m=1}^{M} \langle A_n A_m \rangle\, \phi_n(x_i)\, \phi_m(x_i)$$

$$= \sum_{n=1}^{M} \sum_{m=1}^{M} \delta_{nm} \lambda_n\, \phi_n(x_i) \phi_m(x_i)$$

$$R_{ii} = \sum_{n=1}^{M} \lambda_n \phi_n^2(x_i)$$

summing up the diagonal elements gives the total variance

$$\sum_{i=1}^{M} R_{ii} = \sum_{i=1}^{M} \sum_{n=1}^{M} \lambda_n \phi_n^2(x_i)$$

$$= \sum_{n=1}^{M} \lambda_n \sum_{i=1}^{M} \phi_n^2(x_i)$$

using (1) $\sum_{i=1}^{M} \phi_n(x_i)\phi_m(x_i) = \delta_{mn}$

$$= \sum_{n=1}^{M} \lambda_n$$

Hence the trace T of the matrix R_{ij}

$$T = \sum_{i=1}^{M} R_{ii} = \sum_{n=1}^{M} \lambda_n = \sum_{n=1}^{M} \langle A_n^2 \rangle$$

using (2)

and we can thus interpret the eigenvalues λ_n as the variances associated with the n-th mode because the new orthogonal "variables" $A_n(t)$ can be viewed as amplitudes whose sum of square equals the trace of the covariance matrix R_{ij} which is the sum of the variances of the individual records $u_i(t)$, i.e,

$$\frac{\lambda_n}{\sum_{n=1}^{M} \lambda_n}$$ is the % variance explained by the n-th mode